GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 19, 2002, 07:51:16 ; Search time 199 Seconds
(without alignments)
3951.492 Million cell updates/sec

Title: US-09-807-459-2
Perfect score: 2359
Sequence: 1 MAPSDSVGDVTKTLLAASES.....DPSKALIRKVSTEADNLLEK 458

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09807459/runat_18102002_141111_28571/app_query.fasta_1.647
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09807459_@CGN_1_1_48_@runat_18102002_141111_28571 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2359	100.0	1828	22	AAF59961	Babesia caballi me
	2	996.5	42.2	1542	14	AAQ47074	21B4/rhoptry genes
	3	827.5	35.1	1962	16	AAQ90252	Babesia merozoite
	4	826.5	35.0	1990	14	AAQ33064	Encodes Babesia bo
	5	826.5	35.0	1990	17	AAT18995	Babesia merozoite
	6	770.5	32.7	1371	14	AAQ47076	B. canis 21B4/rhop
	7	759.5	32.2	1491	14	AAQ47075	21B4/rhoptry antig
	8	161	6.8	282	13	AAQ26065	21B4 gene clone pT
	9	147	6.2	1664976	19	AAV21209	Methanococcus jann
	10	138	5.8	170	13	AAQ26066	21B4 gene clone pT
c	11	134	5.7	26776	20	AAX20254	Borrelia burgdorfe
c	12	122.5	5.2	111309	20	AAX20250	Borrelia burgdorfe
c	13	122.5	5.2	910715	20	AAX20248	Borrelia burgdorfe
	14	121	5.1	11165	21	AAA15186	DNA encoding Esche
	15	120	5.1	3883	20	AAV71915	S. cerevisiae CIN8
	16	120	5.1	3884	22	AAH78010	Nucleotide sequenc
	17	118.5	5.0	4766	15	AAQ70102	Malarial PfEMP3 ep
	18	118.5	5.0	7326	21	AAA70095	Plasmodium falcipa
c	19	117	5.0	33303	21	AAA81514	N. meningitidis pa
	20	117	5.0	349980	21	AAF21610	Neisseria meningit
	21	115	4.9	6637	23	AAS71042	DNA encoding novel
	22	113.5	4.8	4443	23	ABL04821	Drosophila melanog
	23	113.5	4.8	10478	20	AAV84691	Arabidopsis ESD4 (
	24	113	4.8	4282	23	AAS76165	DNA encoding novel
	25	113	4.8	11087	23	AAS74637	DNA encoding novel
	26	112.5	4.8	2375	22	AAK68993	Human immune/haema
	27	112.5	4.8	42738	22	AAK68992	Human immune/haema
	28	111	4.7	1422	20	AAX35720	cDNA encoding a pr
	29	111	4.7	1613	20	AAX35721	cDNA encoding a pr
	30	111	4.7	3278	18	AAT73285	K. lactis origin o
	31	111	4.7	3278	18	AAT62358	Kluyveromyces lact
	32	110	4.7	1777	11	AAQ06842	RNA segment 4 codi
	33	109.5	4.6	1816	23	AAS74418	DNA encoding novel
	34	109.5	4.6	2277	23	AAS69641	DNA encoding novel
	35	109.5	4.6	2277	23	AAS71279	DNA encoding novel
	36	109.5	4.6	2277	23	AAS74306	DNA encoding novel
	37	109.5	4.6	2277	23	AAS74426	DNA encoding novel
	38	109.5	4.6	2277	23	AAS74599	DNA encoding novel
	39	109.5	4.6	2277	23	AAS79126	DNA encoding novel
	40	109.5	4.6	2394	23	AAS84093	DNA encoding novel
	41	109.5	4.6	5569	23	AAS78798	DNA encoding novel
c	42	109.5	4.6	1664976	19	AAV21209	Methanococcus jann
	43	109	4.6	1572	20	AAX61708	B. burgdorferi ant
	44	109	4.6	2746	23	AAS82185	DNA encoding novel
	45	109	4.6	3165	23	AAS78781	DNA encoding novel

ALIGNMENTS

RESULT 1
AAF59961
ID AAF59961 standard; cDNA; 1828 BP.
XX
AC AAF59961;
XX
DT 22-MAY-2001 (first entry)
XX
DE Babesia caballi merozoite 48 kD rhoptry protein-encoding cDNA.
XX
KW Merozoite protein; 48 kD rhoptry protein; antigen; antibody;
KW recombinant production; diagnosis; equine babesiosis;
KW parasitic infection; veterinary; ss.
XX
OS Babesia caballi.
XX
PN WO200112813-A1.
XX

PD 22-FEB-2001.
XX
PF 13-AUG-1999; 99WO-JP04386.
XX
PR 13-AUG-1999; 99WO-JP04386.
XX
PA (KAGA) CHEMO-SERO-THERAPEUTIC RES INST.
PA (MIKA/) MIKAMI T.
XX
PI Mikami T, Ikadai H, Igarashi I, Suzuki N, Nagasawa H, Fujisaki K;
XX
DR WPI; 2001-202867/20.
DR P-PSDB; AAB60669.
XX
PT Gene encoding merozoite protein of Babesia caballi for diagnosis of
PT equine babesiosis caused by this organism -
XX
PS Claim 3; Page 19-22; 27pp; Japanese.
XX
CC The invention relates to a 48 kD merozoite rhoptry protein from Babesia
CC caballi (AAB60669) and cDNA encoding it (AAF59961). The invention also
CC relates to phage vectors containing a nucleic acid encoding the
CC merozoite protein, a method for the recombinant production of the
CC protein, an antibody against the protein, and a method for the diagnosis
CC of equine babesiosis from horse blood samples by using the antibody to
CC detect Babesia caballi merozoites, or by using the 48 kD protein as an
CC antigen to detect anti-Babesia caballi antibodies. The 48 kD merozoite
CC protein, or an antibody specific for the protein may be used for the
CC diagnosis of equine babesiosis caused by Babesia caballi. The present
CC sequence represents cDNA encoding the Babesia caballi merozoite 48 kD
CC rhoptry protein.
XX
SQ Sequence 1828 BP; 523 A; 412 C; 460 G; 433 T; 0 other;

Alignment Scores:

Pred. No.:	1.46e-222	Length:	1828
Score:	2359.00	Matches:	458
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0

US-09-807-459-2 (1-458) x AAF59961 (1-1828)

```
Qy      1 MetAlaProSerAspSerValGlyAspValThrLysThrLeuLeuAlaAlaSerGluSer 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     39 ATGGCTCCCAGCGACTCTGTGGGCGACGTGACTAAGACCTTATTGGCTGCCAGCGAAAGT 98

Qy     21 ValAspSerAlaAlaAsnAlaTyrMetIleAsnSerAspMetSerAspTyrLeuSerAla 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     99 GTGGACTCAGCTGCCAATGCCTATATGATCAACAGTGACATGAGCGATTACTTGTCGGCT 158

Qy     41 ValSerAspAsnPheAlaGluArgIleCysSerGlnValProLysGlySerAsnCysSer 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159 GTGTCTGACAACTTCGCCGAGCGCATTTGCAGTCAGGTCCCTAAGGGGAGTAACTGCAGT 218

Qy     61 AlaSerValSerAlaTyrMetSerArgCysAlaLysGlnAspCysLeuThrLeuGlnSer 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219 GCTTCCGTTAGCGCATACATGAGTCGCTGCGCTAAACAGGACTGCCTGACTCTCCAAAGT 278

Qy     81 LeuLysTyrProLeuGluAlaLysTyrGlnProLeuThrLeuProAspProTyrGlnLeu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279 CTTAAGTACCCTCTTGAGGCTAAGTACCAACCGCTGACCCTTCCTGACCCCTACCAGTTG 338

Qy    101 GluAlaAlaPheIleLeuPheLysGluSerAspAlaAsnProAlaAsnSerThrGluLys 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 GAGGCCGCATTTATACTCTTCAAGGAGAGTGACGCTAATCCGGCCAATAGCACTGAGAAG 398

Qy    121 ArgPheTrpMetArgPheArgArgGlyLysAsnHisSerTyrPheHisAspLeuValPhe 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    399 CGCTTCTGGATGCGTTTCAGAAGGGGCAAGAACCACAGTTACTTCCACGACTTAGTCTTC 458

Qy    141 AsnLeuLeuGluLysAsnValThrArgAspAlaAspAlaThrAspIleGluAsnPheAla 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 AATCTGCTGGAGAAGAACGTGACTCGCGACGCGGATGCTACTGACATTGAGAACTTTGCG 518

Qy    161 SerArgTyrLeuTyrMetAlaThrLeuTyrTyrLysThrTyrThrAsnValAspGluPhe 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    519 TCCAGGTACCTGTACATGGCCACGCTTTACTACAAGACGTACACGAATGTTGATGAGTTC 578

Qy    181 GlyAlaSerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLys 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 GGTGCTAGCTTCTTTAACAAGTTGTCTTTCACTACTGGGTTGTTCGGCTGGGGCATCAAG 638

Qy    201 ArgAlaLeuLysGlnIleIleArgSerAsnLeuProLeuAspIleGlyThrGluHisSer 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    639 AGGGCACTTAAGCAGATTATTCGCTCTAACCTGCCCCTTGACATCGGGACAGAACACAGC 698

Qy    221 ValSerArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIlePro 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    699 GTCAGTCGCCTGCAGCACATTACGAGCAGTTACAAGGATTACATGGATACGCAGATTCCT 758

Qy    241 AlaLeuProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThr 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    759 GCACTGCCCAAGTTTGCGAAGCGTTTCTCCCTTATGGTAGTGCAGAGGCTGCTGGCCACC 818

Qy    261 ValAlaGlyTyrValAspThrProTrpTyrLysLysTrpTyrMetLysLeuLysAsnPhe 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    819 GTGGCTGGTTACGTCGACACCCCGTGGTATAAGAAGTGGTACATGAAGCTGAAGAACTTT 878

Qy    281 MetValAsnArgValPheIleProThrLysLysPhePheAsnLysGluIleArgGluPro 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    879 ATGGTGAACAGGGTGTTCATTCCTACAAAGAAGTTCTTCAATAAGGAAATTCGTGAGCCT 938

Qy    301 SerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGluAsnLysIle 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    939 AGTAAGGCATTAAAAGAAAAGGTGTCAACCGACACCAAGGATTTATTCGAGAACAAAATT 998

Qy    321 GlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspProSerLysAlaLeuLys 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    999 GGGCAGGGTACTGTGGACTTCTTCAATAAGGAAATTCGTGACCCTAGTAAGGCATTAAAA 1058

Qy    341 GluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThrVal 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1059 GAAAAAGTGTCAAACGACGCCAAGGATTTATTCGAGAACAAAATTGGGCAGGGTACTGTG 1118

Qy    361 AspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValSerThr 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1119 GACTTCATCAATAACGAAATTCGTGACCCTAGTAAGGCATTAATAAGAAAAGTGTCAACG 1178

Qy    381 GlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnAsn 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1179 GGGGCCGAGGATTTATTCGAGAACAAAATTGGGCAGGGTACTGTGGACTTCATCAATAAC 1238

Qy    401 GluIleArgAspProSerLysAlaLeuIleArgLysValTyrThrGluAlaAspAspLeu 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1239 GAAATTCGTGACCCTAGTAAGGCATTAATAAGAAAAGTGTACACCGAGGCCGATGATTTA 1298

Qy    421 PheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnLysGluIleArgAspPro 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1299 TTCGAGAACAAAATTGGGCAGGGTACTGTGGACTTCATCAATAAGGAAATTCGTGACCCT 1358

Qy    441 SerLysAlaLeuIleArgLysValSerThrGluAlaAspAsnLeuLeuGluLys 458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1359 AGTAAGGCATTAATAAGAAAAGTGTCTACCGAGGCCGATAATTTATTGGAGAAA 1412
```

RESULT 2
AAQ47074
ID AAQ47074 standard; DNA; 1542 BP.
XX
AC AAQ47074;
XX
DT 13-JAN-1994 (first entry)
XX
DE 21B4/rhoptry genes 1-4 representative DNA.
XX

KW Polymerase chain reaction; PCR; amplify; primer; detection;
KW babesiosis; parasite; Babesia bovis; 21B4/rhoptry; antigen; gene;
KW repeat region; immune response; vaccine; ss.
XX
OS Synthetic.
XX
PN WO9314204-A.
XX
PD 22-JUL-1993.
XX
PF 15-JAN-1993; 93WO-AU00012.
XX
PR 15-JAN-1992; 92AU-0000399.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX
PI Dalrymple BP, Peters JM;
XX
DR WPI; 1993-243219/30.
DR P-PSDB; AAR39900.
XX
PT Detecting closely linked gene copies which encode protective
PT antigen against babesiosis - by screening babesial genomic DNA
PT library with oligo-nucleotide probe based partial sequencing of
PT protective antigen and identifying positive clones
XX
PS Claim 22; Fig 5; 55pp; English.
XX
CC This sequence is a gene which is representative of the Babesia bovis
CC 21B4/rhoptry antigen gene region. This sequence was isolated by PCR
CC using the primers given in AAQ47068-72. Primer 21B4.1 corresponds to
CC part of the repeated region of 21B4/rhoptry antigen. In hybridisation
CC assays this primer recognised two tandemly repeated regions suggesting
CC that B. bovis contains two copies of the 21B4/rhoptry antigen gene.
CC The two proteins encoded by the two antigen genes are identical.
CC Primers 21B4.2 and 21B4.3 flank the 21B4-309 coding region of the
CC antigen gene. Primer 21B4.4 primes synthesis just 3' to the end of
CC the open reading frame. The entire open reading frame was shown to
CC encode five antigen genes. The 3' non-repetitive sequences of open
CC reading frames 1-4 are identical. Gene 5 shows sequence divergence
CC throughout most of the open reading frame. Babesia antigen genes can
CC be used in the production of a combined vaccine which will stimulate
CC a greater immune response and afford broader immunity than a single
CC antigen vaccine.
XX
SQ Sequence 1542 BP; 455 A; 366 C; 343 G; 378 T; 0 other;

Alignment Scores:

Pred. No.:	1.67e-88	Length:	1542
Score:	996.50	Matches:	204
Percent Similarity:	62.64%	Conservative:	81
Best Local Similarity:	44.84%	Mismatches:	153
Query Match:	42.24%	Indels:	18
DB:	14	Gaps:	6

US-09-807-459-2 (1-458) x AAQ47074 (1-1542)

```
Qy       1 MetAlaProSerAspSerValGlyAspValThrLysThrLeuLeuAlaAlaSerGluSer 20
           :::||||||::::::   ||||||:::|||      ||||||      |||   |||
Db      91 CTAGCACCTGCTGAAGTAGTCGGTAATGTGGCCGCTACTTTGTCTACCGCAGATGAGATT 150

Qy      21 ValAspSerAlaAlaAsnAlaTyrMetIleAsnSerAspMetSerAspTyrLeuSerAla 40
           :::   |||                  :::   :::||||||   |||   :::
Db     151 ATATCATCACACGACTATGATAAGGAGTTGATAAATGATATGAGAGATCAGATGAGGGAA 210

Qy      41 ValSerAspAsnPheAlaGluArgIleCysSerGlnValProLysGlySerAsnCysSer 60
           :::   :::||||||   :::   :::|||::::::   |||:::   |||||||||
Db     211 ATGGGTGAGAACTTCGTTGACACCGTATGTACGAAGGCCCCTGAAGATTCCAACTGTCGC 270

Qy      61 AlaSerValSerAlaTyrMetSerArgCysAlaLysGlnAspCysLeuThrLeuGlnSer 80
             |||:::   |||      ||||||            ||||||   :::   :::
Db     271 CAAATGGTTGCCCTCTATGCTGACCGTTGTGAGATGTACGGTTGTTTGCAGATTGACAAC 330

Qy      81 LeuLysTyrProLeuGluAlaLysTyrGlnProLeuThrLeuProAspProTyrGlnLeu 100
           :::   ||||||::::::   :::||||||||||||:::||||||:::||||||||||||
Db     331 GTCAACTACCCCGTTGATGAGGAATACCAACCCTTATCACTCCCCAACCCCTACCAATTG 390

Qy     101 GluAlaAlaPheIleLeuPheLysGluSerAspAlaAsnProAlaAsnSerThrGluLys 120
           :::||||||||||   |||||||||   |||   :::|||||||||   :::      |||
Db     391 GATGCTGCCTTCACTTTGTTCAAGAACTCTGCGTCGAACCCAGCTAAAAATGGTTTAAAA 450

Qy     121 ArgPheTrpMetArgPheArgArgGlyLysAsnHisSerTyrPheHisAspLeuValPhe 140
             ||||||||||||||||   ||||||   |||      :::|||      :::
Db     451 GGTCAATGGATGCGTTTCCGAAACGGGAAAAAACATGGTGACTACCACCGTTTCCTCATT 510

Qy     141 AsnLeuLeuGluLysAsnValThrArgAspAlaAspAlaThrAspIleGluAsnPheAla 160
             ||||||   |||   |||||||||          |||||||||:::|||
Db     511 GGCTTGTTGGCAAAAGGCGTTACTCGTAAGGATGGTGCTACTGACCTTGAGTTCCTCGTC 570

Qy     161 SerArgTyrLeuTyrMetAlaThrLeuTyrTyrLysThrTyrThrAsnValAspGluPhe 180
           ::::::   ||||||||||||||||   ||||||||||||||||      |||:::
Db     571 AACAAGCTTTTGTACATGGCCACCACCTATTACAAAACCTACCTAATTGTCAACAACAGC 630

Qy     181 GlyAlaSerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLys 200
             |||   |||||||||:::::::::|||   |||   :::||||||:::|||||||||
Db     631 GATGCCAAATTCTTTAACAGGATTGCTTTCGCAACAAAGATTTTCGGTTTTGGTATCAAG 690

Qy     201 ArgAlaLeuLysGlnIleIleArgSerAsnLeuProLeuAspIleGlyThrGluHisSer 220
           :::||||||||||   |||:::||||||||:::|||      :::|||   |||||||||
Db     691 AAGGCGCTCAAGGATATCGTCCGCTCCAATGTTCCCGAATATATGGGA---GAGCACAGT 747

Qy     221 ValSerArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIlePro 240
           :::   |||:::   |||:::          ||||||||||||||   ||||||:::|||
Db     748 ATTGAACGCATTAGTCATTTGGCCCACGGTTACAAGGACTACATGTTGACCCAAGTTCCA 807

Qy     241 AlaLeuProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThr 260
             |||   |||||||||:::|||:::|||   ||||||::::::   ||||||::::::
Db     808 ACCCTTTCAAAGTTTGCCGAACGTTACTCTGACATGGTTATGAAAGTTCTGTTGAGCAGC 867

Qy     261 ValAlaGlyTyrValAspThrProTrpTyrLysLysTrpTyrMetLysLeuLysAsnPhe 280
           :::||||||||||||      ||||||||||||:::|||:::   |||:::|||:::|||
Db     868 CTCGCCGGTTATGTCAAGGCTCCATGGTACAAGAGATGGTTCAACAAGGTTAAAGATTTC 927

Qy     281 MetValAsnArgValPheIleProThrLysLysPheAsnLysGluIleArgGluPro 300
           ||||||::::::      |||||||||:::   |||:::|||      :::   |||
Db     928 TTTGTTAACAAAATAGGCAAACCTACAAAGGAACACTTCCACAAG------AAACACCCA 981

Qy     301 SerLysAla---LeuLysGluLysValSerThrAspThrLysAspLeuPheGluAsnLys 319
             |||         :::|||:::         |||||||||   ||||||||||||
Db     982 CGCACCGCCGAATTTTTCGACAAGATGCATGAACGCACCAAGGACTTCTTCGAGAACAAG 1041

Qy     320 IleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspProSerLysAlaLeu 339
           ||||||      |||   |||||||||      :::|||      |||:::|||
Db    1042 ATAGGTGCTCCTACCAAGGACTTCTTCGAGAACAAGATAGGTGCTCCTACCAAGGACTTC 1101

Qy     340 LysGlu---LysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGlnGly 358
             |||   |||:::            ||||||   ||||||||||||||||||
Db    1102 TTCGAGAACAAGATAGGTGCTCCTACCAAGGACTTCTTCGAGAACAAGATAGGTGCTCCT 1161

Qy     359 ThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysVal 378
           |||   ||||||      |||:::|||      |||:::|||
Db    1162 ACCAAGGACTTCTTCGAGAACAAGATAGGTGCTCCTACCAAA---------------- 1203

Qy     379 SerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIle 398
                     |||   ||||||||||||||||||      |||   ||||||
Db    1204 --------------GACTTCTTTGAGAACAAGATAGGTGCTCCTACCAAGGACTTCTTC 1248

Qy     399 AsnAsnGluIleArgAspProSerLysAlaLeuIle---ArgLysValTyrThrGluAla 417
           |||::::::   :::   :::|||                |||:::
Db    1249 GAGAACAAGCTGCCTGAACGCACCAAGGACTTCTTCGAGAACAAGCTGCCTGAACACACC 1308

Qy     418 AspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnLysGluIle 437
           |||   ||||||||||||:::   :::   |||   ||||||         :::|||
Db    1309 AAGGACTTCTTCGAGAACAAGCTGCCTGAACACACCAAGGACTTCTTTGAGAACAAGATT 1368

Qy     438 ArgAspProSerLysAlaLeuIleArgLysValSerThrGluAla 452
```

```
Db   1369 GGTGCCCCCATTAAAGA-ATCATTTGAGAACTTGGGACACAAGCG 1412

RESULT 3
AAQ90252
ID   AAQ90252 standard; cDNA; 1962 BP.
XX
AC   AAQ90252;
XX
DT   14-NOV-1995  (first entry)
XX
DE   Babesia merozoite p58 cDNA.
XX
KW   Merozoite; surface protein; antigen; p58; babesiosis; vaccine; ss.
XX
OS   Babesia bigemina.
XX
FH   Key             Location/Qualifiers
FT   CDS             186..1628
FT                   /*tag= a
FT   sig_peptide     186..248
FT                   /*tag= b
FT   mat_peptide     249..1631
FT                   /*tag= c
FT   polyA_signal    1884..1889
FT                   /*tag= d
XX
PN   US5422428-A.
XX
PD   06-JUN-1995.
XX
PF   27-MAR-1987;   87US-0031328.
XX
PR   06-DEC-1991;   91US-0803636.
PR   27-MAR-1987;   87US-0031328.
PR   01-MAR-1991;   91US-0663255.
XX
PA   (UNIW ) UNIV WASHINGTON STATE.
XX
PI   Davis WC,  McElwain TF,  McGuire TC,  Perryman LE;
XX
DR   WPI; 1995-214706/28.
DR   P-PSDB; AAR77249.
XX
PT   Babesia merozoite 45 kD surface protein from B. bigemina -  used in
PT   vaccines for the prophylaxis of bovine babesiosis.
XX
PS   Disclosure; Column 31-34; 30pp; English.
XX
CC   Antigenic surface proteins (45, 55 and 58 kDa) were isolated from
CC   the intraerythrocytic merozoite stage of B. bigemina JG-29.  The 58
CC   kDa surface protein (AAR77249) was characterized, and encoding
CC   cDNA (AAQ90252) was isolated from a lambda GEM11 library.
XX
SQ   Sequence 1962 BP; 506 A; 442 C; 492 G; 522 T; 0 other;

Alignment Scores:
Pred. No.:              1e-71          Length:         1962
Score:                  827.50         Matches:        181
Percent Similarity:     59.12%         Conservative:   75
Best Local Similarity:  41.80%         Mismatches:     140
Query Match:            35.08%         Indels:         37
DB:                     16             Gaps:           12

US-09-807-459-2 (1-458) x AAQ90252 (1-1962)

Qy      4 SerAspSerValGlyAspValThrLysThrLeuLeuAlaAlaSerGluSerValAspSer 23
Db    282 GCAGAGGTGGTTGGAGATGTGTCCAAGACCTTGCTGGAAGCCAATGAGGTTGTCAATGCT 341

Qy     24 AlaAlaAsnAlaTyrMetIleAsnSerAspMetSerAspTyrLeuSerAlaValSerAsp 43
Db    342 GAAATGGAAGCAACTCAGGTCAACAAAGATATGCAAAGTCAATTGTCTAATGTTAAGGAG 401

Qy     44 AsnPheAlaGluArgIleCysSerGlnValProLysGlySerAsnCysSerAlaSerVal 63
Db    402 ACCATTGTTGGTGAGGTCTGCGAGAAAGTTGCTGGAAACTCTACCTGCGGTGAGAGCGTA 461

Qy     64 SerAlaTyrMetSerArgCysAlaLysGlnAspCysLeuThrLeuGlnSerLeuLysTyr 83
Db    462 ATTGCCTATGTTAACCGTTGTGATGAGGGCGATTGTCTGACGCTTGACAGCATG------ 515

Qy     84 ProLeuGluAlaLysTyrGlnProLeuThrLeuProAspProTyrGlnLeuGluAlaAla 103
Db    516 -----------AAGTACAAGCCGTTGAGTCTGCCAAATCCTTACCAGTTGGACGCTGCC 563

Qy    104 PheIleLeuPheLysGluSerAspAlaAsnProAlaAsnSerThrGluLysArgPheTrp 123
Db    564 TTCATGCTTTTCAGGGAAAGTGATTCTAACCCTGCGAAGAATGAGGTGAAGCGCTTCTGG 623

Qy    124 MetArgPheArgArgGlyLysAsnHisSerTyrPheHisAspLeuValPheAsnLeuLeu 143
Db    624 ATGCGTTCGAGG------AGCAGCCACGGCGACTACCATCACTTTGTTGTTAGCTTGTTG 677

Qy    144 GluLysAsnValThrArgAspAlaAspAlaThrAspIleGluAsnPheAlaSerArgTyr 163
Db    678 AAGAAGAATGTTGTACGCGACCCTGAATCCAATGATGTTGAGAACTTCGCATCGCAGTAC 737

Qy    164 LeuTyrMetAlaThrLeuTyrTyrLysThrTyrThrAsnValAspGluPheGlyAlaSer 183
Db    738 TTCTACATGACTACGTTGTACTACAAGACTTACCTGACCGTTGACTTTACGGCGGCTAAG 797

Qy    184 PhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAlaLeu 203
Db    798 TTCTTCAACAAGCTTGCTTTCACAACTCGCCTGTTCGGTTTCGGTATCCAGAAGGCGTTG 857

Qy    204 LysGlnIleIleArgSerAsnLeuProLeuAspIleGlyThrGluHisSerValSerArg 223
Db    858 AAGCGTTTGGTTAGGAGCAACCTTCCCGTTGACCTTGGAACC---CACCCTGAGGCCACC 914

Qy    224 LeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaLeuPro 243
Db    915 ATCCGCGAAATAGCTAGCGGCTACGGCGAGTACATGATGACCCAGGTGCCTGCGATGACC 974

Qy    244 LysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValAlaGly 263
Db    975 TCGTTCGCTGAGCGTTTCTCCAAGATGGCTACTAAGACTCTGTTGGTTACCGTCAGCGAC 1034

Qy    264 TyrValAspThrProTrpTyrLysLysTrpTyrMetLysLeuLysAsnPheMetValAsn 283
Db   1035 TACGTCCATTTGCCCGCGTACAAGAGGTGGTACAGGAAGTTCAAGGAATTCATTGTGAAC 1094

Qy    284 ArgValPheIleProThrLysLysPhePheAsnLysGluIleArgGluProSerLys--- 302
Db   1095 ---TTCTTTACTGACCCTGCCAAGTTGATTATGAAGCACGTCTCTCAGCCTGTAAAGACT 1151

Qy    303 AlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGluAsnLysIleGlyGln 322
Db   1152 GCCTACACAAAGCTGGTCCCCGAAGAGCACAGGCAGGCTATCAGGAATGTCGTCGGTCAA 1211

Qy    323 GlyThrValAspPhePheAsnLysGluIleArgAspProSerLysAlaLeuLysGluLys 342
Db   1212 AGCACCAAGCATATTGCCAAC---GGTGTACGTGATTTGTCAAGGATGATTAAGGAG--- 1265

Qy    343 ValSerAsnAspAlaLysAspLeuPheGluAsnLysIle-----------------Gly 356
Db   1266 ---------CCTAGCCAACAAATAATTCGTGAGAAGCTGCCTCACTACCTTTCTAAGGCA 1316

Qy    357 GlnGlyThrValAspPheIleAsnAsnGluIleArgAsp------ProSerLysAlaLeu 374
Db   1317 AAGGGAGCCGTTGAGCACGTTGTTAAGAAGGTTAAATCCGTTGTGCCGATAAAG------ 1370

Qy    375 IleArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThr 394
Db   1371 ---CAAAAGGGCGACCAACCATCCGAAGCAGCTGTAGAGGAAACCGTTCCGTCTGGC--- 1424
```

```
Qy     395 ValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValTyr 414
Db    1425 ---GATTCCGCGGAAACTGAATTTGAGGTCCCTGAAGAACAATACGTCGATGCTGTTACT 1481

Qy     415 Thr------------------------GluAlaAspAsp 419
Db    1482 ACTCAGGAGGTTAACAGCGAGAAGGTTGATGCCGACGAT 1520

RESULT 4
AAQ33064
ID   AAQ33064 standard; DNA; 1990 BP.
XX
AC   AAQ33064;
XX
DT   06-MAY-1993  (first entry)
XX
DE   Encodes Babesia bovis 60kD immunoreactive merozoite surface epitope.
XX
KW   babesiosis; cows; cattle; bos taurus; babesia bovis; babesia bigemina;
KW   merozoite; schizont; ss.
XX
OS   Babesia bovis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1990
FT                   /*tag=  a
XX
PN   US5171685-A.
XX
PD   15-DEC-1992.
XX
PF   04-APR-1990;   90US-0504461.
XX
PR   04-APR-1990;   90US-0504461.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (USDA ) US SEC OF AGRIC.
XX
PI   Davis WC,  Goff WL,  Hines SA,  Jasmer DP,  Mcelwain TF;
PI   Mcguire TC,  Palmergh,  Perryman LE,  Reduker DW;
XX
DR   WPI; 1993-008582/01.
DR   P-PSDB; AAR30613.
XX
PT   DNA encoding Babesia bovis protein - is used as probes and for
PT   prodn. of polypeptide(s) for use in vaccines and for prodn. of
PT   antibodies
XX
PS   Example 19; Fig 3; 20pp; English.
XX
CC   This sequence encodes an immunoreactive epitope located on the
CC   surface of babesia bovis merozoites. A. B. bovis cDNA expression
CC   library was constructed using poly A(+) RNA isolated from B. bovis
CC   infected blood cultures. Plaques were immunoscreened using rabbit
CC   anti-Bv60 sera, and positive plaques tested for reactivity with
CC   monoclonal antibodies that recognised a Bv42 surface exposed epitope
CC   as well as an isotype control monoclonal antibody and normal rabbit
CC   serum. Lambda rBv60 phagemid DNA was isolated from bacteria, and
CC   then restriction digested.
XX
SQ   Sequence 1990 BP; 628 A; 437 C; 398 G; 527 T; 0 other;

Alignment Scores:
Pred. No.:              1.28e-71      Length:          1990
Score:                  826.50        Matches:         178
Percent Similarity:     51.30%        Conservative:    79
Best Local Similarity:  35.53%        Mismatches:      195
Query Match:            35.04%        Indels:          49
DB:                     14            Gaps:            6

US-09-807-459-2 (1-458) x AAQ33064 (1-1990)

Qy       1 MetAlaProSerAspSerValGlyAspValThrLysThrLeuLeuAlaAlaSerGluSer 20

Db     212 CTCGCTCCAGCTGAAGTGGTAGGTGATTTAACCTCCACATTGGAAACAGCTGATACTTTG 271

Qy      21 ValAspSerAlaAlaAsnAlaTyrMetIleAsnSerAspMetSerAspTyrLeuSerAla 40
Db     272 ATGACTCTCCGTGACCACATGCACAACATTACTAAGGATATGAAACATGTTTTGAGCAAT 331

Qy      41 ValSerAspAsnPheAlaGluArgIleCysSerGlnValProLysGlySerAsnCysSer 60
Db     332 GGTCGTGAGCAGATTGTAAATGATGTTTGCTCTAATGCTCCTGAGGACTCCAACTGTCGT 391

Qy      61 AlaSerValSerAlaTyrMetSerArgCysAlaLysGlnAspCysLeuThrLeuGlnSer 80
Db     392 GAGGTAGTTAACAATTATGCTGACCGTTGTGAAATGTACGGATGCTTTACGATTGACAAT 451

Qy      81 LeuLysTyrProLeuGluAlaLysTyrGlnProLeuThrLeuProAspProTyrGlnLeu 100
Db     452 GTCAAATATCCGTTGTATCAAGAGTACCAACCTCTATCTCTTCCAAACCCTTACCAGTTG 511

Qy     101 GluAlaAlaPheIleLeuPheLysGluSerAspAlaAsnProAlaAsnSerThrGluLys 120
Db     512 GATGCTGCGTTCAGATTGTTCAAAGAGAGTGCATCGAACCCTGCCAAGAACAGCGTAAAA 571

Qy     121 ArgPheTrpMetArgPheArgArgGlyLysAsnHisSerTyrPheHisAspLeuValPhe 140
Db     572 CGCGAATGGTTGCGTTTCAGAAATGGAGCGAACCATGGTGATTACCACTACTTCGTCACT 631

Qy     141 AsnLeuLeuGluLysAsnValThrArgAspAlaAspAlaThrAspIleGluAsnPheAla 160
Db     632 GGTCTGTTGAACAACAATGTTGTGCACGAGGAAGGAACTACCGATGTTGAATATCTTGTC 691

Qy     161 SerArgTyrLeuTyrMetAlaThrLeuTyrTyrLysThrTyrThrAsnValAspGluPhe 180
Db     692 AACAAGGTACTCTATATGGCTACCATGAACTACAAGACTTATTTGACAGTAAACAGTATG 751

Qy     181 GlyAlaSerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLys 200
Db     752 AACGCCAAGTTCTTCAACAGATTCAGCTTCACTACAAAGATATTCAGTCGTCGTATTAGG 811

Qy     201 ArgAlaLeuLysGlnIleIleArgSerAsnLeuProLeuAspIleGlyThrGluHisSer 220
Db     812 CAAACATTGAGTGATATCATCAGGTGGAATGTTCCTGAAGATTTT---GAAGAAAGGAGC 868

Qy     221 ValSerArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIlePro 240
Db     869 ATCGAACGTATCACTCAACTTACTAGCAGCTACGAAGATTACATGTTGACCCAGATTCCA 928

Qy     241 AlaLeuProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThr 260
Db     929 ACTCTTTCCAAGTTTGCACGTCGTTATGCTGACATGGTGAAGAAGGTTCTGCTCGGTAGC 988

Qy     261 ValAlaGlyTyrValAspThrProTrpTyrLysLysTrpTyrMetLysLeuLysAsnPhe 280
Db     989 TTGACCTCGTACGTTGAAGCTCCTTGGTACAAAAGATGGATAAAGAAATTCAGAGACTTT 1048

Qy     281 MetValAsnArgValPheIleProThrLysLysPhe---------------------- 292
Db    1049 TTCTCTAAAAACGTTACCCAACCTACAAAGAAGTTCATCGAGGATACTAACGAAGTTACC 1108

Qy     292 ------------------------------------------------------------ 292
Db    1109 AAAAACTATCTGAAAGCCAATGTTGCTGAGCCCACTAAAAAGTTTATGCAGGACACTCAC 1168

Qy     293 ------------------PheAsnLysGluIleArgGluProSerLysAlaLeuLysGlu 306
Db    1169 GAAAAAACCAAAGGCTATCTGAAAGAGAATGTAGCCGAACCTACTAAGACTTTTTTCAAG 1228

Qy     307 LysValSerThrAspThrLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAsp 326
Db    1229 GAGGCTCCTCAAGTCACCAAACACTTCTTCGATGAGAACATTGGCCAACCCACCAAGGAG 1288

Qy     327 PhePheAsnLysGluIleArgAspProSerLysAlaLeuLysGluLysValSerAsnAsp 346
```

```
Db   1289 TTTTTCAGGGAAGCTCCCCAAGCCACTAAACATTTCCTAGACGAAAACATCGGTCAACCA 1348

Qy    347 AlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGlu 366

Db   1349 ACCAAGGAGTTCTTC---AGGGAGGCTCCTCAAGCCACTAAGCACTTCCTAGGCGAGAAT 1405

Qy    367 IleArgAspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPhe 386

Db   1406 ATTGCTCAACCTACTAAAGAATTTTTCAAGGATGTCCCTCAAGTCACCAAGAAGGTTATA 1465

Qy    387 GluAsnLysIleGlyGlnGlyThrValAspPhe-------------------------- 397

Db   1466 ACTGAGAACATTGCTCAACCAACTAAGGAGTTCCGGAGGGAGGTTCCTCATGCTACCATG 1525

Qy    398 ------IleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValTyrThr 415

Db   1526 AAAGTCTTGAATGAAAACATTGCTCAACCTGCCAAGGAAATCATACATGAGTTTGGTACA 1585

Qy    416 GluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnLys 435

Db   1586 GGCGCCAAGAAT---TTCATTTCCGCAGCCCATGAAGGTACTAAGCAGTTCTTAAACGAA 1642

Qy    436 GluIleArgAspProSerLysAlaLeuIleArg---LysValSerThrGluAlaAspAsn 454

Db   1643 ACTGTTGGCCAACCTACAAAGGAATTCCTGAACGGAGCTTTAGAAACTACTAAAGACGCA 1702

Qy    455 Leu 455

Db   1703 TTA 1705
```

```
RESULT 5
AAT18995
ID   AAT18995 standard; cDNA; 1990 BP.
XX
AC   AAT18995;
XX
DT   15-OCT-1996  (first entry)
XX
DE   Babesia merozoite surface protein cDNA clone Bv60.
XX
KW   Babesiosis; merozoite protein; vaccine; probe; diagnosis; ss.
XX
OS   Babesia bovis.
XX
FH   Key             Location/Qualifiers
FT   CDS             122..1819
FT                   /*tag= a
XX
PN   US5518916-A.
XX
PD   21-MAY-1996.
XX
PF   04-APR-1989;   89US-0333155.
XX
PR   04-APR-1990;   90US-0504461.
PR   04-APR-1989;   89US-0333155.
PR   14-DEC-1992;   92US-0989616.
PR   21-NOV-1994;   94US-0342480.
XX
PA   (USDA ) US SEC OF AGRIC.
XX
PI   Goff WL,  Jasmer DP,  McElwain TF,  McGuire TC,  Reduker DW;
PI   Stiller D;
XX
DR   WPI; 1996-259067/26.
DR   P-PSDB; AAR97981.
XX
PT   New fragment of Babesia bovis genomic DNA - useful as a probe for
PT   detecting Babesia infection
XX
PS   Example 18; Fig 3; 19pp; English.
XX
CC   A cDNA clone (AAT18995) codes for Bv60, a 60 kDa immunoreactive
CC   protein (AAR97981) located on the surface of Babesia bovis merozoites.
CC   It was isolated from a blood-stage B. bovis cDNA library in lambda
CC   ZAPII by subcloning into Bluescript SK(-) and immunoscreening using
CC   monospecific anti-Bv60 antisera.  Bv60, Bo44 and Bv42 (see also
CC   AAT18993 and AAT18994) DNA sequences can be used to make recombinant
CC   proteins useful as vaccines for the prophylaxis of bovine babesiosis.
CC   They can also be used as diagnostic probes.
XX
SQ   Sequence 1990 BP; 628 A; 437 C; 398 G; 527 T; 0 other;
```

```
Alignment Scores:
Pred. No.:              1.28e-71      Length:          1990
Score:                  826.50        Matches:         178
Percent Similarity:     51.30%        Conservative:    79
Best Local Similarity:  35.53%        Mismatches:      195
Query Match:            35.04%        Indels:          49
DB:                     17            Gaps:            6

US-09-807-459-2 (1-458) x AAT18995 (1-1990)

Qy      1 MetAlaProSerAspSerValGlyAspValThrLysThrLeuLeuAlaAlaSerGluSer 20

Db    212 CTCGCTCCAGCTGAAGTGGTAGGTGATTTAACCTCCACATTGGAAACAGCTGATACTTTG 271

Qy     21 ValAspSerAlaAlaAsnAlaTyrMetIleAsnSerAspMetSerAspTyrLeuSerAla 40

Db    272 ATGACTCTCCGTGACCACATGCACAACATTACTAAGGATATGAAACATGTTTTGAGCAAT 331

Qy     41 ValSerAspAsnPheAlaGluArgIleCysSerGlnValProLysGlySerAsnCysSer 60

Db    332 GGTCGTGAGCAGATTGTAAATGATGTTTGCTCTAATGCTCCTGAGGACTCCAACTGTCGT 391

Qy     61 AlaSerValSerAlaTyrMetSerArgCysAlaLysGlnAspCysLeuThrLeuGlnSer 80

Db    392 GAGGTAGTTAACAATTATGCTGACCGTTGTGAAATGTACGGATGCTTTACGATTGACAAT 451

Qy     81 LeuLysTyrProLeuGluAlaLysTyrGlnProLeuThrLeuProAspProTyrGlnLeu 100

Db    452 GTCAAATATCCGTTGTATCAAGAGTACCAACCTCTATCTCTTCCAAACCCTTACCAGTTG 511

Qy    101 GluAlaAlaPheIleLeuPheLysGluSerAspAlaAsnProAlaAsnSerThrGluLys 120

Db    512 GATGCTGCGTTCAGATTGTTCAAAGAGAGTGCATCGAACCCTGCCAAGAACAGCGTAAAA 571

Qy    121 ArgPheTrpMetArgPheArgArgGlyLysAsnHisSerTyrPheHisAspLeuValPhe 140

Db    572 CGCGAATGGTTGCGTTTCAGAAATGGAGCGAACCATGGTGATTACCACTACTTCGTCACT 631

Qy    141 AsnLeuLeuGluLysAsnValThrArgAspAlaAspAlaThrAspIleGluAsnPheAla 160

Db    632 GGTCTGTTGAACAACAATGTTGTGCACGAGGAAGGAACTACCGATGTTGAATATCTTGTC 691

Qy    161 SerArgTyrLeuTyrMetAlaThrLeuTyrTyrLysThrTyrThrAsnValAspGluPhe 180

Db    692 AACAAGGTACTCTATATGGCTACCATGAACTACAAGACTTATTTGACAGTAAACAGTATG 751

Qy    181 GlyAlaSerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLys 200

Db    752 AACGCCAAGTTCTTCAACAGATTCAGCTTCACTACAAAGATATTCAGTCGTCGTATTAGG 811

Qy    201 ArgAlaLeuLysGlnIleIleArgSerAsnLeuProLeuAspIleGlyThrGluHisSer 220

Db    812 CAAACATTGAGTGATATCATCAGGTGGAATGTTCCTGAAGATTTT---GAAGAAAGGAGC 868

Qy    221 ValSerArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIlePro 240

Db    869 ATCGAACGTATCACTCAACTTACTAGCAGCTACGAAGATTACATGTTGACCCAGATTCCA 928

Qy    241 AlaLeuProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThr 260

Db    929 ACTCTTTCCAAGTTTGCACGTCGTTATGCTGACATGGTGAAGAAGGTTCTGCTCGGTAGC 988

Qy    261 ValAlaGlyTyrValAspThrProTrpTyrLysLysTrpTyrMetLysLeuLysAsnPhe 280
```

```
Db    989 TTGACCTCGTACGTTGAAGCTCCTTGGTACAAAAGATGGATAAAGAAATTCAGAGACTTT 1048

Qy    281 MetValAsnArgValPheIleProThrLysLysPhe---------------------- 292
Db   1049 TTCTCTAAAAACGTTACCCAACCTACAAAGAAGTTCATCGAGGATACTAACGAAGTTACC 1108

Qy    292 ------------------------------------------------------------ 292
Db   1109 AAAAACTATCTGAAAGCCAATGTTGCTGAGCCCACTAAAAAGTTTATGCAGGACACTCAC 1168

Qy    293 ------------------PheAsnLysGluIleArgGluProSerLysAlaLeuLysGlu 306
Db   1169 GAAAAAACCAAAGGCTATCTGAAAGAGAATGTAGCCGAACCTACTAAGACTTTTTTCAAG 1228

Qy    307 LysValSerThrAspThrLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAsp 326
Db   1229 GAGGCTCCTCAAGTCACCAAACACTTCTTCGATGAGAACATTGGCCAACCCACCAAGGAG 1288

Qy    327 PhePheAsnLysGluIleArgAspProSerLysAlaLeuLysGluLysValSerAsnAsp 346
Db   1289 TTTTTCAGGGAAGCTCCCCAAGCCACTAAACATTTCCTAGACGAAAACATCGGTCAACCA 1348

Qy    347 AlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGlu 366
Db   1349 ACCAAGGAGTTCTTC---AGGGAGGCTCCTCAAGCCACTAAGCACTTCCTAGGCGAGAAT 1405

Qy    367 IleArgAspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPhe 386
Db   1406 ATTGCTCAACCTACTAAAGAATTTTTCAAGGATGTCCCTCAAGTCACCAAGAAGGTTATA 1465

Qy    387 GluAsnLysIleGlyGlnGlyThrValAspPhe--------------------------- 397
Db   1466 ACTGAGAACATTGCTCAACCAACTAAGGAGTTCCGGAGGGAGGTTCCTCATGCTACCATG 1525

Qy    398 ------IleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValTyrThr 415
Db   1526 AAAGTCTTGAATGAAAACATTGCTCAACCTGCCAAGGAAATCATACATGAGTTTGGTACA 1585

Qy    416 GluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnLys 435
Db   1586 GGCGCCAAGAAT---TTCATTTCCGCAGCCCATGAAGGTACTAAGCAGTTCTTAAACGAA 1642

Qy    436 GluIleArgAspProSerLysAlaLeuIleArg---LysValSerThrGluAlaAspAsn 454
Db   1643 ACTGTTGGCCAACCTACAAAGGAATTCCTGAACGGAGCTTTAGAAACTACTAAAGACGCA 1702

Qy    455 Leu 455
Db   1703 TTA 1705
```

RESULT 6
AAQ47076
ID AAQ47076 standard; DNA; 1371 BP.
XX
AC AAQ47076;
XX
DT 13-JAN-1994 (first entry)
XX
DE B. canis 21B4/rhoptry antigen gene 2 DNA.
XX
KW Polymerase chain reaction; PCR; amplify; primer; detection;
KW babesiosis; parasite; Babesia bovis; 21B4/rhoptry; antigen; gene;
KW repeat region; immune response; vaccine; ss.
XX
OS Babesia canis.
XX
PN WO9314204-A.
XX
PD 22-JUL-1993.
XX
PF 15-JAN-1993; 93WO-AU00012.
XX
PR 15-JAN-1992; 92AU-0000399.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX
PI Dalrymple BP, Peters JM;
XX
DR WPI; 1993-243219/30.
DR P-PSDB; AAR39902.
XX
PT Detecting closely linked gene copies which encode protective
PT antigen against babesiosis - by screening babesial genomic DNA
PT library with oligo-nucleotide probe based partial sequencing of
PT protective antigen and identifying positive clones
XX
PS Claim 24; Fig 7; 55pp; English.
XX
CC This sequence represents the Babesia canis 21B4/rhoptry antigen gene
CC 2. This sequence was determined from restriction fragments from the
CC clone B. canis lambda GEM-11 #9. B. canis was found to contain two
CC genes which are related to the B. bovis 21B4 gene. Gene 1 and gene 2
CC are very similar but gene 2 appears to contain a large number of
CC repeats. Babesia antigen genes can be used in the production of a
CC combined vaccine which will stimulate a greater immune response and
CC afford broader immunity than a single antigen vaccine. See also
CC AAQ47068-74.
XX
SQ Sequence 1371 BP; 418 A; 312 C; 325 G; 318 T; 0 other;

Alignment Scores:

Pred. No.:	2.55e-66	Length:	1371
Score:	770.50	Matches:	161
Percent Similarity:	55.65%	Conservative:	95
Best Local Similarity:	35.00%	Mismatches:	161
Query Match:	32.66%	Indels:	43
DB:	14	Gaps:	9

US-09-807-459-2 (1-458) x AAQ47076 (1-1371)

```
Qy      1 MetAlaProSerAspSerValGlyAspValThrLysThrLeuLeuAlaAlaSerGluSer 20
Db     91 CTTTCTAAATCAGATGGAGCCGAAAAGACCTTGTCTACCCTTCTTAACGTTGACGCATCG 150

Qy     21 ValAspSerAlaAlaAsnAlaTyrMetIleAsnSerAspMetSerAspTyrLeuSerAla 40
Db    151 ACGAGAGCGGCTTTAGAAGGCTACAGAATGAATGCTGCTATGGCAAACTTTTCAAACGGT 210

Qy     41 ValSerAspAsnPheAlaGluArgIleCysSerGlnValProLysGlySerAsnCysSer 60
Db    211 CGCAGGGAAGAGGAGGAGGAAGCTGTCTGTGGAAACATCGCAGAGGAGACTGAATGTCAG 270

Qy     61 AlaSerValSerAlaTyrMetSerArgCysAlaLysGlnAspCysLeuThrLeuGlnSer 80
Db    271 AAGAGCGTAGCTGAGTATGTTGAAAGCTGCGTCAGGTACGACTGCTTTAGCATTGAGAAC 330

Qy     81 LeuLysTyrProLeuGluAlaLysTyrGlnProLeuThrLeuProAspProTyrGlnLeu 100
Db    331 CAGAAGTACCCTCAAGAGAAGGAATACCAGCCGCTGACCCTTCCCAACCCTTATCAATTG 390

Qy    101 GluAlaAlaPheIleLeuPheLysGluSerAspAlaAsnProAlaAsnSerThrGluLys 120
Db    391 GAGGCCGCGTTCTATGTCTTCAGAAACAGTGAATCAAACCCAATTAAAAACCCAACGGAA 450

Qy    121 ArgPheTrpMetArgPheArgArgGlyLysAsnHisSerTyrPheHisAspLeuValPhe 140
Db    451 GCTTTCTGGATGCGTTTTCGTCACGGCGGACGTTATGGCGCCTATCACAATTTCCTGGTG 510

Qy    141 AsnLeuLeuGluLysAsnValThrArgAspAlaAspAlaThrAspIleGluAsnPheAla 160
Db    511 AATATTCTATATAAAAACCTCAGCGATAGCATGGTGGATGATAACCTCGAAGGTTTCGTA 570

Qy    161 SerArgTyrLeuTyrMetAlaThrLeuTyrTyrLysThrTyrThrAsnValAspGluPhe 180
Db    571 CGCAAATATGCTTACATGGCCACCATGTATTACAAGACATACACCGCTTTGGATGTTGTA 630
```

```
Qy    181 GlyAlaSerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLys 200
Db    631 AATGCCAGGATCATTAACAAAATTGCCTTCTCCCGCCACTTGTTCGGAAGGCAGATTAGG 690

Qy    201 ArgAlaLeuLysGlnIleIleArgSerAsnLeuProLeuAspIleGlyThrGluHisSer 220
Db    691 AACGCCCTAACGAATATAATCCGCAGTAACATTCCCGAAGACTTCGGC---AAGTACAAC 747

Qy    221 ValSerArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIlePro 240
Db    748 GTTGACCGCCTGAGGCACGTGATGGGTGGCTACGAGGAGTACATGATGAAGCAAGTACCA 807

Qy    241 AlaLeuProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThr 260
Db    808 TCTCTTCCAAATTTCGCCAAGAAATACGCGGGAATGGTGGTAAAGAGTTTAATTAAGAAT 867

Qy    261 ValAlaGlyTyrValAspThrProTrpTyrLysLysTrpTyrMetLysLeuLysAsnPhe 280
Db    868 GTGGGAGCTTACCAGAAGCAGCCATGGTTCAAAAAGCTGAACAATCAAATAAGAAACTTC 927

Qy    281 MetValAsnArgValPheIleProThrLysLysPhePheAsnLysGluIleArgGluPro 300
Db    928 TTCGTGAACAAGATCCATGAACCCACCAAGGAGTTCTTCGTGAACAAGATCCATGAACCC 987

Qy    301 SerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGluAsnLysIle 320
Db    988 ----------------------------------ACCAAGGAGTTCTTCGTGAACAAGATC 1014

Qy    321 GlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspProSerLysAlaLeuLys 340
Db   1015 CATGAACCCACCAAGGAGTTCTTCGTGAACAAGATCCATGAACCCACC------------ 1062

Qy    341 GluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThrVal 360
Db   1063 --------------------AAGGAGTTCTTCGTGAACAAGCTCCATGAACCCACCAAG 1101

Qy    361 AspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValSerThr 380
Db   1102 GAGTTCTTCGTGAACAAGCTCCATGAACCCACCAAGGAGTTCTTCTCTAACATGGTACCC 1161

Qy    381 GlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnAsn 400
Db   1162 GGGGCATTCCAGAAGATATCTGAAAAGGCAGGTAGA---------------------- 1197

Qy    401 GluIleArgAspProSerLysAlaLeuIleArgLysValTyrThrGluAlaAspAspLeu 420
Db   1198 CACCTGAGAAGT---TCTAAGACAGTTGTCCCTGAA------GACGAGCCCTCATCGTCT 1248

Qy    421 PheGluAsnLysIle---------GlyGlnGlyThrVal---AspPheIleAsnLysGlu 436
Db   1249 CTGGAAAATGAGGCTGTAGAGGATGGTCAACTTACAATGGGGGATGTGACCGATTTTGAA 1308

Qy    437 IleArgAspPro--------------SerLysAlaLeuIleArgLysValSerThrGlu 451
Db   1309 ATGGCGACTCCCACCTATGAGCAGGGCTCACAGGAGAGTTTAAATGAAGTTGGTAACGAA 1368
```

RESULT 7
AAQ47075
ID AAQ47075 standard; DNA; 1491 BP.
XX
AC AAQ47075;
XX
DT 13-JAN-1994 (first entry)
XX
DE 21B4/rhoptry antigen gene 5 DNA.
XX
KW Polymerase chain reaction; PCR; amplify; primer; detection;
KW babesiosis; parasite; Babesia bovis; 21B4/rhoptry; antigen; gene;
KW repeat region; immune response; vaccine; ss.
XX
OS Synthetic.
XX
PN WO9314204-A.
XX
PD 22-JUL-1993.
XX
PF 15-JAN-1993; 93WO-AU00012.
XX
PR 15-JAN-1992; 92AU-0000399.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX
PI Dalrymple BP, Peters JM;
XX
DR WPI; 1993-243219/30.
DR P-PSDB; AAR39901.
XX
PT Detecting closely linked gene copies which encode protective
PT antigen against babesiosis - by screening babesial genomic DNA
PT library with oligo-nucleotide probe based partial sequencing of
PT protective antigen and identifying positive clones
XX
PS Claim 23; Fig 6; 55pp; English.
XX
CC This sequence represents the Babesia bovis 21B4/rhoptry antigen gene
CC 5. This sequence was isolated by PCR using the primers given in
CC AAQ47068-72. Primer 21B4.1 corresponds to part of the repeated
CC region of 21B4/rhoptry antigen. In hybridisation assays this primer
CC recognised two tandemly repeated regions suggesting that B. bovis
CC contains two copies of the 21B4/rhoptry antigen gene. The two
CC proteins encoded by the two antigen genes are identical. Primers
CC 21B4.2 and 21B4.3 flank the 21B4-309 coding region of the antigen
CC gene. Primer 21B4.4 primes synthesis just 3' to the end of the open
CC reading frame. The entire open reading frame was shown to encode five
CC antigen genes. The 3' non-repetitive sequences of open reading frames
CC 1-4 are identical. Gene 5 shows sequence divergence throughout most
CC of the open reading frame. Babesia antigen genes can be used in the
CC production of a combined vaccine which will stimulate a greater immune
CC response and afford broader immunity than a single antigen vaccine.
XX
SQ Sequence 1491 BP; 479 A; 286 C; 328 G; 398 T; 0 other;

Alignment Scores:

Pred. No.:	3.45e-65	Length:	1491
Score:	759.50	Matches:	161
Percent Similarity:	54.04%	Conservative:	93
Best Local Similarity:	34.26%	Mismatches:	173
Query Match:	32.20%	Indels:	43
DB:	14	Gaps:	8

US-09-807-459-2 (1-458) x AAQ47075 (1-1491)

```
Qy      1 MetAlaProSerAspSerValGlyAspValThrLysThrLeuLeuAlaAlaSerGluSer 20
Db     91 CTAGCACCTGCTGAAGTAGTCGGTGATCTTACTCATACACTGAAAATTGCTGACGAAATT 150

Qy     21 ValAspSerAlaAlaAsnAlaTyrMetIleAsnSerAspMetSerAspTyrLeuSerAla 40
Db    151 ATAAATGCCGAAAACATAGAACATGAAATAAATCGTGATATGCATTTAAGACTGGTAGAA 210

Qy     41 ValSerAspAsnPheAlaGluArgIleCysSerGlnValProLysGlySerAsnCysSer 60
Db    211 GAGGGAAGTAAATTTATTGATCAAATTTGTCAGGAAGTCGCTGAGGATTCTAAGTGCCGT 270

Qy     61 AlaSerValSerAlaTyrMetSerArgCysAlaLysGlnAspCysLeuThrLeuGlnSer 80
Db    271 GAACAAGTTGAATCATATGTTAAGCGTTGTGAAGAGAACAACTGTTTACAGATTGACGAA 330

Qy     81 LeuLysTyrProLeuGluAlaLysTyrGlnProLeuThrLeuProAspProTyrGlnLeu 100
Db    331 GTAGCTTACCCTTTGAATCAGGAATATCAGCCACTATTACTTCCGGAACCATACCAATTG 390

Qy    101 GluAlaAlaPheIleLeuPheLysGluSerAspAlaAsnProAlaAsnSerThrGluLys 120
Db    391 GATGCTGCCTTCACATTGTTCAAGAACTGTGAGTCTAATCCTGCTAAAAATGGCCTGAAA 450
```

```
Qy    121 ArgPheTrpMetArgPheArgArgGlyLysAsnHisSerTyrPheHisAspLeuValPhe 140
Db    451 GGTCCATGGATGCGTTACAAGGAGGGTAAAGAGCATGGTGACTACCACCATTTCATCATA 510

Qy    141 AsnLeuLeuGluLysAsnValThrArgAspAlaAspAlaThrAspIleGluAsnPheAla 160
Db    511 AGTTTGCTCGGCAAGAGTTTGGTTCGTAAGGATGGTGTTACTGACCTTGAGTTCCTCGTC 570

Qy    161 SerArgTyrLeuTyrMetAlaThrLeuTyrTyrLysThrTyrThrAsnValAspGluPhe 180
Db    571 AACAAGCTTTTGTACATGGCCACCACCTACTACAAAACCTACCTAATTGTCAAGAAGTTC 630

Qy    181 GlyAlaSerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLys 200
Db    631 GGTGCTCGCTTTTTCAACACTTTCTCTTTTACAATGAATATATTCGGTATAGGTATTAAA 690

Qy    201 ArgAlaLeuLysGlnIleIleArgSerAsnLeuProLeuAspIleGlyThrGluHisSer 220
Db    691 AGGGCATTAAAAGGCATCGTCCGCTCCAATGTTCCGGAAGACATGGGA---GAGCACAGT 747

Qy    221 ValSerArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIlePro 240
Db    748 ATTGAACGTATTAGTCATTTGTCCGAAGGATACAAGGACTACATGTTGACACAGGTGCCA 807

Qy    241 AlaLeuProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThr 260
Db    808 ACCCTTTCAAAGTTTGCCGAACGTTACTCTGACATGGTTATGAAAGTTCTGTTGAGCAGC 867

Qy    261 ValAlaGlyTyrValAspThrProTrpTyrLysLysTrpTyrMetLysLeuLysAsnPhe 280
Db    868 CTCGCCGGTTATGTCAAGGCTCCATGGTACAAGAGATGGATCAATAGATTTAAGTCCTTG 927

Qy    281 MetValAsnArg----------------------------------------ValPheIle 287
Db    928 TTAACTGGTGAAGCTTACAACCCTGATGAAGATATACATTTACTTAAACCTATTTTCGTT 987

Qy    288 ProThrLysLysPhePheAsnLysGluIleArgGluProSerLys---AlaLeuLysGlu 306
Db    988 GATACTCCTAGGAATTACATAAAAGATGCACTTAAGCCACTGCGTGATGCTGTAGAGGAA 1047

Qy    307 LysValSerThrAspThrLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAsp 326
Db   1048 AATATCGTTAACCCCGTCAGTGATTATCTCCGACGCAAGCAAAATATTTCTAGGTCACAA 1107

Qy    327 PhePheAsn--------------------LysGluIleArgAspProSerLysAlaLeu 339
Db   1108 AACTACAATGACGGGCATCATAAAATAGATCCCTCTCTTTATGAACCTAAGAGGCCGCAT 1167

Qy    340 LysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThr 359
Db   1168 ATCGGAATTGCTGCGAATCACGCTAGGGATTATATAGATGATAAGGTAAATAAA------ 1221

Qy    360 ValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValSer 379
Db   1222 --------------------------------GCGAAAGAATTAGTGTCCGCTGCCAAG 1248

Qy    380 ThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsn 399
Db   1249 GACCGCGCAACAGGCATATATGCAGATCACGTGAAACCTGCTCTGAGTGATATTACGAAC 1308

Qy    400 ---------AsnGluIleArgAspProSerLysAlaLeuIleArgLysValTyr------ 414
Db   1309 GTTGTAAAGAATGATCTTCTTGATGCAGTAAAC------ATCAGGAATATATTACGTGGC 1362

Qy    415 ---ThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIle 433
Db   1363 AGCTCTCAAGATGACAATAATGAACAAGAGAAGACTGAGGAGGAGAAAGTTGAGGAAGTT 1422

Qy    434 AsnLysGluIleArgAspProSerLysAla 443
Db   1423 AAACCTGAGCTGAAGCAAAAAGAGTATGCT 1452
```

```
RESULT 8
AAQ26065
ID   AAQ26065 standard; DNA; 282 BP.
XX
AC   AAQ26065;
XX
DT   09-DEC-1992  (first entry)
XX
DE   21B4 gene clone pT#13, EcoRI insert.
XX
KW   Beta-galactosidase; B. bovis; Bb; T21B4; ss.
XX
OS   Babesia bovis.
XX
PN   EP492525-A.
XX
PD   01-JUL-1992.
XX
PF   20-DEC-1991;   91EP-0121990.
XX
PR   21-DEC-1990;   90AU-0004051.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Casu RE,  Commins MA;
XX
DR   WPI; 1992-218727/27.
DR   P-PSDB; AAR25187.
XX
PT   Monoclonal antibody to Babesia bovis parasite - used to isolate
PT   antigens for use in vaccines for treating Babesiosis and
PT   providing immunity in cattle
XX
PS   Claim 13; Fig 9; 24pp; English.
XX
CC   The sequences given in AAQ26062 and AAQ26065-7 are portions of the 21B4
CC   gene which were isolated from a B. bovis (Bb) cDNA lambda gt11
CC   library and cloned into pGEM7Zf(+).  The resulting plasmids were
CC   tranformed into E. coli strain JM83.  The inserts were in frame,
CC   when translated, with the vector beta-galactosidase gene.  The fusion
CC   protiens produced by translation of these vectors were recognised by
CC   the monoclonal antibody of the invention, T21B4.  These fusion
CC   antigens could be used in vaccines for the treatment of babesiosis
CC   and to provide immunity in relation to Bb infection in cattle against
CC   different strains of Babesia by heterologous and homologous challenge.
XX
SQ   Sequence 282 BP; 80 A; 60 C; 63 G; 79 T; 0 other;
```

Alignment Scores:			
Pred. No.:	3.18e-07	Length:	282
Score:	161.00	Matches:	32
Percent Similarity:	51.65%	Conservative:	15
Best Local Similarity:	35.16%	Mismatches:	44
Query Match:	6.82%	Indels:	0
DB:	13	Gaps:	0

US-09-807-459-2 (1-458) x AAQ26065 (1-282)

```
Qy      2 AlaProSerAspSerValGlyAspValThrLysThrLeuLeuAlaAlaSerGluSerVal 21
Db     10 GCTCCAGCTGAAGTGGTAGGTGATTTAATCTCCACATTGGAAACAGCTGATACTTTGATG 69

Qy     22 AspSerAlaAlaAsnAlaTyrMetIleAsnSerAspMetSerAspTyrLeuSerAlaVal 41
Db     70 ACTCTCCGTGACCACATGCACAACATTACTAAGGATATGAAACACGTTTTGAGCAATGGT 129

Qy     42 SerAspAsnPheAlaGluArgIleCysSerGlnValProLysGlySerAsnCysSerAla 61
Db    130 CGTGAGCAGATTGTAAATGATGTTTGCTCTAATGCTCCTGAGGACTCCAACTGTCGTGAG 189

Qy     62 SerValSerAlaTyrMetSerArgCysAlaLysGlnAspCysLeuThrLeuGlnSerLeu 81
Db    190 GTAGTTAACAACTATGCTGACCGTTGTGAAATGTACGGATGCTTCACGATTGACAATGTC 249
```

```
Qy      82 LysTyrProLeuGluAlaLysTyrGlnProLeu 92
           :::||||||||||      :::|||||||||||||
Db     250 AGATATCCGTTGTACCAAGAGTACCAACCTCTA 282

RESULT 9
AAV21209
ID   AAV21209 standard; DNA; 1664976 BP.
XX
AC   AAV21209;
XX
DT   10-NOV-1998  (first entry)
XX
DE   Methanococcus jannaschii circular chromosome.
XX
KW   Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KW   genome; autotrophic; extrachromosomal element; identification; ds.
XX
OS   Methanococcus jannaschii.
XX
PN   WO9807830-A2.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-US14900.
XX
PR   22-AUG-1996;   96US-0024428.
XX
PA   (GENO-) INST GENOMIC RES.
PA   (UNII ) UNIV ILLINOIS FOUND.
PA   (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI   Bult CJ,  Smith HO,  Venter JC,  White OR,  Woese CR;
XX
DR   WPI; 1998-169145/15.
XX
PT   Complete genome sequence of methano-genic archaeon, Methanoccoccus
PT   jannaschii  - useful in identification of M. jannaschii genome
PT   fragment
XX
PS   Claim 13; Page 152-585; 614pp; English.
XX
CC   The present sequence represents the complete 1.66-megabase pair genome
CC   sequence of the Methanococcus jannaschii circular chromosome. The
CC   present invention describes M. jannaschii open reading frames from the
CC   genome sequence. The invention also describes a computer based system
CC   for identifying fragments of the M. jannaschii genome that are
CC   homologous to target nucleotide sequences, comprising: (a) data storage
CC   means comprising the nucleotide sequence of the 1664976, 58407 or 16550
CC   bp sequence (see AAV21209, AAV21210 and AAV21211), or a nucleotide
CC   sequence at least 99.9% identical to it; (b) search means for comparing a
CC   target sequence to the nucleotide sequence of the data storage means to
CC   identify a homologous sequence, and (c) retrieval means for obtaining
CC   the homologous sequence. The method, which is based on whole genome
CC   random sequencing of an autotrophic archaeon M. jannaschii, the genome
CC   of which consists of 3 physically distinct elements, a large circular
CC   chromosome (the 1664976 bp sequence given in AAV21209), a large circular
CC   extra-chromosomal element (the 58407 bp sequence given in AAV21210), and
CC   a small circular extra-chromosomal element (the 16550 bp sequence given
CC   in AAV21211), can be used in the identification of M. jannaschii genome
CC   fragment.
XX
SQ   Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;

Alignment Scores:
Pred. No.:              0.823          Length:          1664976
Score:                  147.00         Matches:         88
Percent Similarity:     38.79%         Conservative:    59
Best Local Similarity:  23.22%         Mismatches:      150
Query Match:            6.23%          Indels:          82
DB:                     19             Gaps:            17

US-09-807-459-2 (1-458) x AAV21209 (1-1664976)
```

```
Qy     151 AlaAspAlaThrAspIleGlu-AsnPheAlaSerArgTyrLeuTyrMetAla-------- 167
           |||   |||   |||:::||| ||||||   :::|||||||   |||
Db 1418187 GCCAAAGCCCTTGATTTAGAAAAATTTTTTTAACAGATATTGTTACAAAACTAAAAACAA 1418246

Qy     168 ------------ThrLeuTyrTyrLysThrTyrThrAsnValAspGluPheGlyAlaSe 183
                        :::   :::      ||||||         |||      |||   ||
Db 1418247 TAACTGCTATATATCAGCATTTTTAATAACATATTTATACCCAGAC---ATGGGATTAAG 1418303

Qy     183 rPhePheAsnLysLeuSerPheThrThrGlyLeuPhe---------GlyTrpGlyIleLy 200
           ||||      :::              |||||||||         |||   |||
Db 1418304 TTTTAATGGTGAAAACATGATTATCTTCGGATTATTTGGAAAAACAGGATGTGGAAAGAC 1418363

Qy     200 sArgAlaLeuLysGlnIleIleArgSerAsnLeuProLeuAspIle-----------Gl 216
                 |||   ::::::   :::   :::      :::|||||||
Db 1418364 AGAAATATTAAATGAATTAAAGAAACATCATCCTGTAATAGATATTGAAGAAATTGCAAG 1418423

Qy     216 yThrGluHisSerVal---SerArgLeuGlnHisIleThrSerSerTyrLysAspTyrMe 235
           |||      |||:::         |||   |||::::::         :::
Db 1418424 AACAAGAGGGAGTATTTTAGGGGATTTATATCACTTAAGTATGAGAAGCCAGGGGGAGTT 1418483

Qy     235 tAspThrGlnIleProAlaLeuProLysPheAlaLysArgPheSerLeuMetValValGl 255
           |||      |||          :::   ||||||:::|||        ||||||::
Db 1418484 TGACTATCTAATAAATAAAGAAATTGAAAAGGCTAAAAAATTTGGTTATGCAGTGGTTGA 1418543

Qy     255 n-----------------------------------------ArgLeuLeuAlaTh 260
           :                                         |||||||||
Db 1418544 ATATGAGGGAAGGAAGATTGGTGGAGAGAAAAAGCTAAAAATTCCTGAGTTGTTGGCTGA 1418603

Qy     260 rValAlaGlyTyr--------------ValAspThrProTrp---------------- 269
           :::      |||             :::|||   |||:::
Db 1418604 TATTAAAAACTATACTTACAAAATCTTAATTGACTGTCCTTATGAATGCCAGATAAACAG 1418663

Qy     270 ------------TyrLysLysTrpTyrMetLysLeuLysAsnPheMetValAsnArgVa 285
                       ||||||          |||   |||      ::::::|||:::
Db 1418664 ATTAGTCTCCATTTATAAGCCAAAAAATGAAAAAGAGAAAGAAATTTTGATAAACAAATT 1418723

Qy     285 lPheIleProThrLysLysPheAsnLysGluIleArgGluProSerLysAlaLeuLy 305
             |||      :::   |||         |||:::   |||           :::
Db 1418724 TTTAATATTAAAGGAGAGTTTTAAAAAGCCAGAGATGATTGAAGCAGTTGATAACATCAT 1418783

Qy     305 sGluLysValSerThrAsp-----------ThrLysAspLeuPheGluAsnLysIleGl 321
           |||   :::      |||                |||   |||   |||:::
Db 1418784 TGAACTCATAAAACAAGACAAATACTATGAAGCAGCAAAATTAATTGAAGAAAAGCTTTA 1418843

Qy     321 y--------------------------------GlnGlyThrValAspPhePheAsnLy 330
                                           :::      :::      :::|||::
Db 1418844 TAGAGAACATTATATGAGAAATGTGAAAAAGATAAAGCCAGATTTAATTGTTTATAATGA 1418903

Qy     330 sGluIleArgAspProSerLysAlaLeuLysGluLysValSerAsnAspAlaLysAsp-- 349
           ::::::::::      :::|||   :::   |||   :::          |||:::
Db 1418904 GGATGTTAAAAAATCAGCTAAAATAATTGATGAATTTATTAAGAAAAAATTAAAGCAGCA 1418963

Qy     350 -------LeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIl 367
                  :::   |||   |||   |||:::|||              ||||||
Db 1418964 TAATTTAATTTAAGAGGGAAAACATGGACGAGGGAATTTTAGCTCGGCTAATAACCTTTA 1419023

Qy     367 eArg-AspProSerLysAlaLeuIle-----------ArgLysValSerThrGlyAlaG 383
           ||| |||              ::::::           ||||||:::   ||||||   :
Db 1419024 CAGAGGATGTTGTTTTATGTATTGTTTTAAATGATGGGAGGAAGATGATAACTGGTGGTA 1419083

Qy     383 luAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIle---------AsnA 400
           ::   :::         ||||||                ||||||       :::
Db 1419084 AAAAAATATTGGCTGGAAAAATTGAAGGAGAGCTTGCTTCTTTTATATTATCTGCCTCTA 1419143

Qy     400 snGluIleArgAspProSerLysAlaLeuIleArgLysValTyrThrGluAlaAspAspL 420
            |||      :::      |||      :::::::|||   :::   :::   |||
Db 1419144 AAGAATTTTTAGAAGATAAAAAGGTCGGTGTAAAAAAA---TTTAAAGATTACGACATAT 1419200

Qy     420 euPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnLysGluIleArgAspP 440
           ||||||                ::::|||   |||||||||   ::::::
Db 1419201 ACTTTGAAAGA----------------ATAGAT---ATCAACAAGTTTTTAAAATCCA 1419239
```

```
Qy    440 roSerLysAlaLeuIleArgLysValSerThrGluAlaAspAsnLeuLeuGlu 457
                          ::::::         |||   ::::::   |||||||||
Db 1419240 TTGGGGGAGAATTTGTTAAAAATACAATAACTGTTAGTGAG---TTGTTAGAG 1419289
```

RESULT 10
AAQ26066
ID AAQ26066 standard; DNA; 170 BP.
XX
AC AAQ26066;
XX
DT 09-DEC-1992 (first entry)
XX
DE 21B4 gene clone pT#13, EcoRI insert (2).
XX
KW Beta-galactosidase; B. bovis; Bb; T21B4; ss.
XX
OS Babesia bovis.
XX
PN EP492525-A.
XX
PD 01-JUL-1992.
XX
PF 20-DEC-1991; 91EP-0121990.
XX
PR 21-DEC-1990; 90AU-0004051.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX
PI Casu RE, Commins MA;
XX
DR WPI; 1992-218727/27.
DR P-PSDB; AAR25188.
XX
PT Monoclonal antibody to Babesia bovis parasite - used to isolate
PT antigens for use in vaccines for treating Babesiosis and
PT providing immunity in cattle
XX
PS Claim 15; Fig 10; 24pp; English.
XX
CC The sequences given in AAQ26062 and AAQ26065-7 are portions of the 21B4
CC gene which were isolated from a B. bovis (Bb) cDNA lambda gt11
CC library and cloned into pGEM7Zf(+). The resulting plasmids were
CC tranformed into E. coli strain JM83. The inserts were in frame,
CC when translated, with the vector beta-galactosidase gene. The fusion
CC protiens produced by translation of these vectors were recognised by
CC the monoclonal antibody of the invention, T21B4. These fusion
CC antigens could be used in vaccines for the treatment of babesiosis
CC and to provide immunity in relation to Bb infection in cattle against
CC different strains of Babesia by heterologous and homologous challenge.
XX
SQ Sequence 170 BP; 56 A; 28 C; 31 G; 51 T; 4 other;

Alignment Scores:

Pred. No.:	2.98e-05	Length:	170
Score:	138.00	Matches:	27
Percent Similarity:	66.67%	Conservative:	9
Best Local Similarity:	50.00%	Mismatches:	18
Query Match:	5.85%	Indels:	0
DB:	13	Gaps:	0

US-09-807-459-2 (1-458) x AAQ26066 (1-170)

```
Qy    161 SerArgTyrLeuTyrMetAlaThrLeuTyrTyrLysThrTyrThrAsnValAspGluPhe 180
          ::::::   ||||||||||||||||:::   ||||||||||||      |||:::
Db      1 AACAAGGTACTTTATATGGCTACCATGGACTACAAGACTTATTTGACAGTAAACAGTATG 60

Qy    181 GlyAlaSerPheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLys 200
            |||   |||||||||:::   ||||||||||||   :::|||         |||:::
Db     61 AACGCCANNTTCTTCAACAGATTCAGCTTTACTACAAAGATATTCAGTNNTCGTATTAGG 120

Qy    201 ArgAlaLeuLysGlnIleIleArgSerAsnLeuProLeuAsp 214
          :::   |||       |||||||||   |||:::|||   |||
Db    121 CAAACATTGAGTGATATCATCAGGTGGAATGTTCCTGAAGAT 162
```

RESULT 11
AAX20254/c
ID AAX20254 standard; DNA; 26776 BP.
XX
AC AAX20254;
XX
DT 04-MAY-1999 (first entry)
XX
DE Borrelia burgdorferi polynucleotide sequence #7.
XX
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
XX
OS Borrelia burgdorferi.
XX
PN WO9858943-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12764.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
PI White OR;
XX
DR WPI; 1999-081217/07.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
XX
PS Claim 1; Page 867-882; 1128pp; English.
XX
CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
XX
SQ Sequence 26776 BP; 8604 A; 4331 C; 4257 G; 9580 T; 4 other;

Alignment Scores:

Pred. No.:	0.0633	Length:	26776
Score:	134.00	Matches:	61
Percent Similarity:	40.96%	Conservative:	75
Best Local Similarity:	18.37%	Mismatches:	124
Query Match:	5.68%	Indels:	72
DB:	20	Gaps:	13

US-09-807-459-2 (1-458) x AAX20254 (1-26776)

```
Qy    167 AlaThrLeuTyrTyrLysThrTyrThrAsnValAspGluPheGlyAlaSerPhePheAsn 186
          ||||||:::|||      |||
Db   6709 GCTACAATATACTTAACTACC----------------------------AAATTGGGTGAG 6677

Qy    187 LysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAlaLeuLysGlnIle 206
            :::::::::      |||   :::      |||:::|||         |||   |||
Db   6676 CTTGTAGCATATCTTGCCGGCAAGTGGGTGCAAGGCATGAAAATGTTGTTCAAGGCCATT 6617

Qy    207 IleArgSerAsnLeuProLeuAspIleGlyThrGluHisSerValSerArgLeuGlnHis 226
```

```
                    :::   |||      |||   ::::::   :::   :::|||:::
Db   6616 ACCACGCCCTTTGTTAAGCTCTACGAGGGAATCAAATACGTAATCGGCAAGCTAAAAGAA 6557

Qy    227 IleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaLeuProLysPheAla 246
          :::   :::            |||:::   :::           |||   :::|||
Db   6556 CTTGCAACA-----------TATGTTTCTAGTGGATTT------TTAGATCGTTTTAAG 6515

Qy    247 LysArgPheSerLeuMetVal------ValGlnArgLeuLeuAlaThrValAlaGlyTyr 264
            |||      ::::::      ||||||:::|||:::      ::::::
Db   6514 GGTGCGTTTGAGCCCATTATCAAATGGGTTCAAAAGTTGATAGATATGATATCC------ 6461

Qy    265 ValAspThrProTrpTyrLysLysTrpTyrMetLysLeuLysAsnPheMetVal------ 282
                          |||   |||   ||||||||||||   :::
Db   6460 --------------------AAAGCGTATACAAAATTGAAAAATCTGGTTACTTTTTGG 6422

Qy    283 ----------------------------------------------AsnArgValPhe 286
                                                           ::::::   |||
Db   6421 AAAAAAGCCGAAAAGGAAAAAGACACTTCTGGTTCCGAACCCGAAAGGGATAAAAAGTTT 6362

Qy    287 IleProThrLysLysPhe---PheAsnLysGluIleArgGluProSerLysAlaLeuLys 305
            |||      |||         ||||||::::::   |||      :::   |||:::
Db   6361 GATCCAAATGCAAAAACCAATATTAACAAGAAAATGGCAGAAGACTATCAAAAGCTGCAA 6302

Qy    306 GluLysValSerThrAspThrLysAspLeuPheGluAsnLysIleGlyGlnGly------ 323
          :::::::::           :::|||::::::   ||||||   |||:::
Db   6301 GACGAAATATTTAATAGGCAGAGGGACATCTAT---AATAAGACCGGGAAAGCACGAGAA 6245

Qy    324 -----------------------ThrValAspPhePheAsnLysGluIleArgAspPro 335
                                  |||::::::      |||::::::   :::|||
Db   6244 CAAGCCTTAAGGAATCTTGAAAAAACCATAAATGAAAAAAACCAAAAGTTTAAGGATGAA 6185

Qy    336 SerLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPhe----------- 351
                :::   ::::::::::::::::::   |||   :::
Db   6184 TACTCCAAAATATTTGATCAGTTAACTGACGAGAACAAGAAAATTTTAGTCGGAGTTGAA 6125

Qy    352 ------GluAsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAsp 369
                |||:::                  ||||||:::|||
Db   6124 AAATCAGTAAATGAGTTCAATAACTCCAATTATGATTTTGTGAATGAG----------- 6077

Qy    370 ProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLys 389
                :::   |||::::::::::      :::        ::::::   :::
Db   6076 ---TACCAAAATTTACTAAAAGAAAAAGAAAGTCGTGAAAGGGAAATAATCAAAACTCTG 6020

Qy    390 IleGlyGlnGlyThrValAspPhe---------IleAsnAsnGluIleArgAspProSer 406
                |||                 :::|||:::||||||   :::   :::
Db   6019 CCCCATACAGATCAAGTAAGCGCTTTACAGAAACTTAATGATGAAATCAATGAAAAGAAC 5960

Qy    407 LysAlaLeuIleArgLysValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGly 426
          ||||||   :::   |||          :::   |||      |||:::
Db   5959 AAAGCGTTCGTAGAAAAATATGGAAAAAGTTTCGAAACTCTG------AACGAGTCTAAC 5906

Qy    427 GlnGlyThrValAspPheIleAsnLysGluIleArgAspProSerLysAlaLeuIleArg 446
          :::      |||      :::   |||:::::::   ::::      |||      :::
Db   5905 AGGCAAGTTGTAGTCGCACTTGAAAAGCAGGTTAATGAATATGAAAAAACCGCTTTAGAT 5846

Qy    447 LysValSerThrGluAlaAspAsnLeuLeuGluLys 458
          :::         ||||||         |||:::|||
Db   5845 CGATCTTTTGTCGAGGCTCAAAAAGCTCTGCAAAAA 5810

RESULT 12
AAX20250/c
ID   AAX20250 standard; DNA; 111309 BP.
XX
AC   AAX20250;
XX
DT   04-MAY-1999  (first entry)
XX
DE   Borrelia burgdorferi polynucleotide sequence #3.
XX
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9858943-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12764.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Clayton R,  Dougherty BA,  Fraser C,  Lathigra R,  Smith HO;
PI   White OR;
XX
DR   WPI; 1999-081217/07.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
XX
PS   Claim 1; Page 738-800; 1128pp; English.
XX
CC   AAX20248 to AAX20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
XX
SQ   Sequence 111309 BP; 35956 A; 13151 C; 19075 G; 43117 T; 10 other;

Alignment Scores:
Pred. No.:              5.75          Length:          111309
Score:                  122.50        Matches:         93
Percent Similarity:     37.47%        Conservative:    55
Best Local Similarity:  23.54%        Mismatches:      144
Query Match:            5.19%         Indels:          103
DB:                     20            Gaps:            20

US-09-807-459-2 (1-458) x AAX20250 (1-111309)

Qy    113 AsnProAlaAsnSerThrGluLysArgPheTrpMetArgPheArgArgGlyLysAsnHis 132
          ||||||   |||                |||
Db  47193 AACCCAATAAACCAAGAAATAAACACATTT---------------------------- 47164

Qy    133 SerTyrPheHisAspLeuValPheAsnLeuLeuGluLysAsnValThrArgAspAlaAsp 152
                |||            |||      |||:::::|||      |||:::
Db  47163 AAAGCTTTTTTAAGTTATGGATTTTATAATCTTAAAAGAAATTTTAACAGAGAA------ 47110

Qy    153 AlaThrAspIleGluAsnPheAlaSerArgTyrLeuTyrMetAlaThrLeuTyrTyrLys 172
                               |||      ||||||      |||   |||      |||
Db  47109 --------------------GCCTTAGAATATTTAAGAAAAGCTTTTCTGATTAAAAAA 47071

Qy    173 ThrTyrThrAsnValAspGluPhe---GlyAlaSerPheAsnLysLeuSerPheThr 191
          |||      :::   :::   :::   |||   :::|||                :::
Db  47070 ACAGATGAAGATGCAAATTATTATCTTGGTGTAGCATTCTTAAAAAACGAAATGTATAAA 47011

Qy    192 ThrGlyLeuPheGlyTrpGlyIleLysArgAlaLeuLysGlnIle-------------- 206
                |||::                  ||||||::::::
Db  47010 GAAGCTCTATAC----------------TATCTTAAAAAAGTCCACAATTTTAATAAA 46969

Qy    207 -----------IleArgSerAsnLeuProLeuAspIleGlyThrGluHisSerValSer 222
                      :::      :::|||   :::   :::            |||
Db  46968 AATAATAAAGATGTCCTAAAACACCTGGGAATAACTTTATTTAATCTAGAAAGCTACAGA 46909
```

```
Qy    223 ArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaLeu 242
          :::         |||   ::::::   |||      :::         |||   ||||||
Db  46908 CAAGCTGTTATAATATTTAACAATATAAAAAAAAATATACAAGGCGACATTGAAGCTCTT 46849

Qy    243 ProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValAla 262
            :::||||||      |||   |||             |||             :::|||
Db  46848 TTGGCATATGCTAAATCTTTATCAAAAATGAATCAAGATCATCTAGCACTAGAGATTGCC 46789

Qy    263 GlyTyrValAspThrProTrpTyrLysLysTrpTyrMetLysLeuLysAsnPheMetVal 282
            :::                           |||   |||:::   |||:::
Db  46788 AACAAAATA---------------------------AAACAAAAAGATGGAATGATT 46759

Qy    283 AsnArgValPheIleProThrLysLysPhePhe------AsnLysGluIleArgGluPro 300
                  :::   |||   :::           |||||||||:::   :::
Db  46758 TATGAGGCTTTATTGATTACATCTGAGATTCATTCAAAAAATAAAGAATTAGAAAAATTA 46699

Qy    301 SerLysAlaLeuLysGlu------LysValSerThrAspThrLys--------------- 313
            :::   :::||||||      ||||||      |||
Db  46698 GAACAAAATATTAAAGAAATAATAAAAGTAAAACCCGACTTGCCTAAAAAAACTTTCCTT 46639

Qy    314 AspLeuPheGluAsnLysIleGlyGln----------------------------Gly 323
          :::||||||   |||   :::|||:::
Db  46638 GAATTATTTTATAAT---CTTGGAGAACTTCAAATCTCTGTTGAAAATTATCAAAAAGCT 46582

Qy    324 ThrValAspPhePheAsnLysGluIleArgAspProSer---LysAlaLeuLysGluLys 342
          |||      |||         |||      ||||||:::   |||   :::||||||||
Db  46581 ACAGAAGCTTTTACCAAAGTTGAAGAGATTGATCCAAATTACAAAAATATTAAAGAAAAA 46522

Qy    343 ValSerAsnAspAlaLysAspLeuPheGluAsn--------------------LysIle 355
          :::         :::|||   |||   ||||||                     :::
Db  46521 TTG---GAATTTAGCAAAAGGTTAAACGAAAACATAGCCCTAAGAATATATTTGCGCAGT 46465

Qy    356 GlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIle 375
          :::         :::   |||   |||||||||                     |||
Db  46464 TCAAAAGAAAATTTTGAAAAAATAGCAAATGAAATT--------------------ATT 46426

Qy    376 ArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrVal 395
            |||:::                  :::   ::::::||||||   :::   |||
Db  46425 TTAAAACTATATTTAAATAAATTTCAAATAAGAGATTCTAAAATAAATGAAGTAACAACA 46366

Qy    396 AspPheIle--------------------AsnAsnGluIleArgAspProSerLysAla 408
            ||||||                     ||||||:::      :::
Db  46365 CAATTTATTGACATAAACTTTCACTTAGCTAACAATCAGTGGGAAGAA------AATTTA 46312

Qy    409 LeuIleArgLysValTyrThrGluAlaAsp---------AspLeuPheGluAsnLysIle 425
          :::::::|||   |||   ||||||   |||         :::||||||
Db  46311 ATAGTGCGTTTTGTAAGAACTGAACAAGATACTTTTGGTGAATTATTC------------ 46264

Qy    426 GlyGlnGlyThrValAspPheIleAsnLysGluIleArgAspProSerLysAlaLeuIle 445
                  |||||||||:::|||         :::      |||   |||
Db  46263 ---------TTAAAAGATTTCATTTCAAAAATAAAAGAAAATAAAATAAAAGGACTCTGT 46213

Qy    446 ---------ArgLysValSerThrGluAlaAspAsnLeuLeuGlu 457
                    |||   |||   :::|||      :::::::|||
Db  46212 ATTGCTCCATCAAAATTCTCTTTAAAGGCTAAACAAATGATTGAA 46168

RESULT 13
AAX20248/c
ID   AAX20248 standard; DNA; 910715 BP.
XX
AC   AAX20248;
XX
DT   04-MAY-1999  (first entry)
XX
DE   Borrelia burgdorferi polynucleotide sequence #1.
XX
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9858943-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12764.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Clayton R,  Dougherty BA,  Fraser C,  Lathigra R,  Smith HO;
PI   White OR;
XX
DR   WPI; 1999-081217/07.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
XX
PS   Claim 1; Page 157-671; 1128pp; English.
XX
CC   AAX20248 to AAX20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
XX
SQ   Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other;

Alignment Scores:
Pred. No.:              95.1          Length:         910715
Score:                  122.50        Matches:        93
Percent Similarity:     37.47%        Conservative:   55
Best Local Similarity:  23.54%        Mismatches:     144
Query Match:            5.19%         Indels:         103
DB:                     20            Gaps:           20

US-09-807-459-2 (1-458) x AAX20248 (1-910715)

Qy    113 AsnProAlaAsnSerThrGluLysArgPheTrpMetArgPheArgArgGlyLysAsnHis 132
          ||||||   |||             |||
Db 636579 AACCCAATAAACCAAGAAATAAACACATTT---------------------------- 636550

Qy    133 SerTyrPheHisAspLeuValPheAsnLeuLeuGluLysAsnValThrArgAspAlaAsp 152
                |||             |||      |||:::::::|||      |||:::
Db 636549 AAAGCTTTTTTAAGTTATGGATTTTATAATCTTAAAAGAAATTTTAACAGAGAA------ 636496

Qy    153 AlaThrAspIleGluAsnPheAlaSerArgTyrLeuTyrMetAlaThrLeuTyrTyrLys 172
                                  |||      ||||||      |||   |||      |||
Db 636495 --------------------GCCTTAGAATATTTAAGAAAAGCTTTTCTGATTAAAAAA 636457

Qy    173 ThrTyrThrAsnValAspGluPhe---GlyAlaSerPhePheAsnLysLeuSerPheThr 191
          |||      :::   :::   :::   |||   :::|||                :::
Db 636456 ACAGATGAAGATGCAAATTATTATCTTGGTGTAGCATTCTTAAAAAACGAAATGTATAAA 636397

Qy    192 ThrGlyLeuPheGlyTrpGlyIleLysArgAlaLeuLysGlnIle-------------- 206
                |||:::                     ||||||::::::
Db 636396 GAAGCTCTATAC-----------------TATCTTAAAAAAGTCCACAATTTTAATAAA 636355

Qy    207 ------------IleArgSerAsnLeuProLeuAspIleGlyThrGluHisSerValSer 222
                       :::     :::|||   :::   :::           |||
Db 636354 AATAATAAAGATGTCCTAAAACACCTGGGAATAACTTTATTTAATCTAGAAAGCTACAGA 636295

Qy    223 ArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaLeu 242
          :::         |||   ::::::   |||      :::         |||   ||||||
Db 636294 CAAGCTGTTATAATATTTAACAATATAAAAAAAAATATACAAGGCGACATTGAAGCTCTT 636235
```

```
Qy    243 ProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValAla 262
Db 636234 TTGGCATATGCTAAATCTTTATCAAAAATGAATCAAGATCATCTAGCACTAGAGATTGCC 636175

Qy    263 GlyTyrValAspThrProTrpTyrLysLysTrpTyrMetLysLeuLysAsnPheMetVal 282
Db 636174 AACAAAATA----------------------------AAACAAAAAGATGGAATGATT 636145

Qy    283 AsnArgValPheIleProThrLysLysPhePhe------AsnLysGluIleArgGluPro 300
Db 636144 TATGAGGCTTTATTGATTACATCTGAGATTCATTCAAAAAATAAAGAATTAGAAAAATTA 636085

Qy    301 SerLysAlaLeuLysGlu------LysValSerThrAspThrLys--------------- 313
Db 636084 GAACAAAATATTAAAGAAATAATAAAAGTAAAACCCGACTTGCCTAAAAAAACTTTCCTT 636025

Qy    314 AspLeuPheGluAsnLysIleGlyGln----------------------------Gly 323
Db 636024 GAATTATTTTATAAT---CTTGGAGAACTTCAAATCTCTGTTGAAAATTATCAAAAAGCT 635968

Qy    324 ThrValAspPhePheAsnLysGluIleArgAspProSer---LysAlaLeuLysGluLys 342
Db 635967 ACAGAAGCTTTTACCAAAGTTGAAGAGATTGATCCAAATTACAAAAATATTAAAGAAAAA 635908

Qy    343 ValSerAsnAspAlaLysAspLeuPheGluAsn--------------------LysIle 355
Db 635907 TTG---GAATTTAGCAAAAGGTTAAACGAAAACATAGCCCTAAGAATATATTTGCGCAGT 635851

Qy    356 GlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIle 375
Db 635850 TCAAAAGAAAATTTTGAAAAAATAGCAAATGAAATT--------------------ATT 635812

Qy    376 ArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrVal 395
Db 635811 TTAAAACTATATTTAAATAAATTTCAAATAAGAGATTCTAAAATAAATGAAGTAACAACA 635752

Qy    396 AspPheIle--------------------AsnAsnGluIleArgAspProSerLysAla 408
Db 635751 CAATTTATTGACATAAACTTTCACTTAGCTAACAATCAGTGGGAAGAA------AATTTA 635698

Qy    409 LeuIleArgLysValTyrThrGluAlaAsp---------AspLeuPheGluAsnLysIle 425
Db 635697 ATAGTGCGTTTTGTAAGAACTGAACAAGATACTTTTGGTGAATTATTC------------ 635650

Qy    426 GlyGlnGlyThrValAspPheIleAsnLysGluIleArgAspProSerLysAlaLeuIle 445
Db 635649 ---------TTAAAAGATTTCATTTCAAAAATAAAAGAAAATAAAATAAAAGGACTCTGT 635599

Qy    446 ---------ArgLysValSerThrGluAlaAspAsnLeuLeuGlu 457
Db 635598 ATTGCTCCATCAAAATTCTCTTTAAAGGCTAAACAAATGATTGAA 635554

RESULT 14
AAA15186
ID   AAA15186 standard; DNA; 11165 BP.
XX
AC   AAA15186;
XX
DT   04-SEP-2000  (first entry)
XX
DE   DNA encoding Escherichia coli virulence proteins.
XX
KW   Virulence protein; tatA; tatB; tatC; tatE; mdoG; creC; recG; yggN;
KW   eck1; iroD; iroC; iroE; mtd2; ms1; vaccine; infection;
KW   Gram negative bacterium; ss.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..1099
FT                   /*tag= a
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93237; no termination codon given"
FT   CDS             1102..1488
FT                   /*tag= b
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93238"
FT   CDS             1573..1896
FT                   /*tag= c
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93239"
FT   CDS             1939..2196
FT                   /*tag= d
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93240; no termination codon given"
FT   CDS             2198..2533
FT                   /*tag= e
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93241"
FT   CDS             2613..3041
FT                   /*tag= f
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93242; no termination codon given"
FT   CDS             3054..3410
FT                   /*tag= g
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93243"
FT   CDS             3460..3705
FT                   /*tag= h
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93244"
FT   CDS             3791..4837
FT                   /*tag= i
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93245"
FT   CDS             4878..7802
FT                   /*tag= j
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93246"
FT   CDS             7816..9483
FT                   /*tag= k
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93247"
FT   CDS             9836..10084
FT                   /*tag= l
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93248"
FT   CDS             10134..10430
FT                   /*tag= m
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93249"
FT   CDS             10459..10779
FT                   /*tag= n
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93250"
XX
PN   WO200028038-A2.
XX
PD   18-MAY-2000.
XX
PF   09-NOV-1999;   99WO-GB03721.
XX
PR   09-NOV-1998;   98GB-0024569.
PR   09-NOV-1998;   98GB-0024570.
PR   17-DEC-1998;   98GB-0027814.
PR   17-DEC-1998;   98GB-0027815.
PR   17-DEC-1998;   98GB-0027816.
PR   17-DEC-1998;   98GB-0027818.
PR   13-JAN-1999;   99GB-0000708.
PR   13-JAN-1999;   99GB-0000710.
PR   13-JAN-1999;   99GB-0000711.
PR   28-JAN-1999;   99GB-0001915.
XX
PA   (MICR-) MICROSCIENCE LTD.
XX
```

```
PI   Crooke HR,  Clarke EE,  Everest PH,  Dougan G,  Holden DW,  Shea JE;
PI   Feldman RG;
XX
DR   WPI; 2000-376550/32.
DR   P-PSDB; AAY93237, AAY93238, AAY93239, AAY93240, AAY93241, AAY93242,
DR   AAY93243.
XX
PT   Peptide encoded by an operon including genes from Escherichia coli for
PT   screening potential drugs, detecting virulence and treating conditions
PT   associated with infection by a Gram negative bacterium -
XX
PS   Disclosure; Page 83-101; 122pp; English.
XX
CC   The present sequence encodes Escherichia coli virulence proteins
CC   The specification describes virulence proteins which are encoded
CC   by an operon including tatA, tatB, tatC, tatE, mdoG, creC, recG, yggN,
CC   eck1, iroD, iroC, iroE, mtd2 or ms1-16 genes obtained from Escherichia
CC   coli K1. The virulence proteins and polynucleotides, and their vaccines
CC   are useful for screening potential drugs, for the detection of virulence,
CC   and for treating or preventing conditions associated with infection by
CC   a Gram negative bacterium particularly Escherichia coli.
XX
SQ   Sequence 11165 BP; 3592 A; 2087 C; 2458 G; 3028 T; 0 other;

Alignment Scores:
Pred. No.:              0.375          Length:          11165
Score:                  121.00         Matches:         89
Percent Similarity:     35.70%         Conservative:    77
Best Local Similarity:  19.14%         Mismatches:      143
Query Match:            5.13%          Indels:          156
DB:                     21             Gaps:            22

US-09-807-459-2 (1-458) x AAA15186 (1-11165)

Qy      99 GlnLeuGluAlaAlaPheIleLeuPheLysGlu---------SerAspAlaAsnProAla 115
           :::||||||   |||   :::   |||              ||||||:::||||||
Db    5604 CGCTTAGAAAAAGCTTTATTAGGTTTTACAAACACAATGTACAGTGATTCAAACCCTCCT 5663

Qy     116 AsnSerThrGluLysArgPheTrpMetArg------------PheArgArgGlyLysAsn 131
                   |||   ::::::                    |||
Db    5664 ATTATAGCTCGTTTTAGAGACTATCTGGAAGATGGTGAGTGCATTGACAGAATTAGCGAA 5723

Qy     132 HisSerTyrPheHisAspLeuValPheAsnLeuLeuGluLysAsnValThrArgAspAla 151
           :::|||           |||||||||   :::   :::
Db    5724 TCAATTTTTTTTACACCGCAAGAATTCAATCTTGCAGATCACCAC--------------- 5768

Qy     152 AspAlaThrAspIleGluAsnPheAlaSerArgTyrLeuTyrMetAlaThrLeuTyrTyr 171
                  ||||||   :::
Db    5769 ------------ATTGAAGGATGG---------------------------------- 5780

Qy     172 LysThrTyrThrAsnValAspGluPheGlyAlaSerPhePheAsnLysLeuSerPheThr 191
                     :::||||||||||               :::   |||   |||
Db    5781 ---------------TTCAATGAATTTGGT--------------CAATTCAGTGGAACT 5810

Qy     192 ThrGlyLeuPheGly--------------------------TrpGlyIleLysArgAla 202
           ::::::|||                             |||
Db    5811 GTTTCTGTTTATGGTGAAGAGCCAATTCATCATGTCGTGACTTGGAAAAATAATAATCAA 5870

Qy     203 LeuLysGln-----------------------------------------IleIleArg 208
           |||   |||                                        :::
Db    5871 TTAACCCAATGCGGTCCATTTAAAATAAAATTAGCGTATATTCATGGTCGGCTTCGTGAT 5930

Qy     209 SerAsnLeuProLeuAspIleGlyThrGluHisSerValSerArgLeuGlnHisIleThr 228
           |||   ||||||:::::::::               :::   |||:::      |||
Db    5931 TCACGCTTACCCATGGAGTTGTGG--------------GCCCCTCTGAAGGAGAAAACA 5975

Qy     229 SerSer-----------------TyrLysAsp-----------------TyrMetAsp 236
                               |||:::|||                |||   |||
Db    5976 GATAGATATGGTGGTTTATATATCTATCGAGATGGATTAAGAATTTTGCCCTATGGAGAT 6035

Qy     237 ThrGlnIleProAlaLeuProLysPheAlaLysArgPheSerLeuMetVal--------- 253
           :::            |||   |||      ||||||   :::|||
Db    6036 TCAGATACGGATTTTCTA---AAAATAGAAAAGAGAAGAACGTTATCCGCTTCTGAATAT 6092

Qy     254 ---------ValGlnArgLeuLeuAlaThrVal-------------------------- 261
                       :::||||||         :::
Db    6093 TTTTTCTCATATCGACGTTTGTTTGGAGCAATAGAATTAACAAAAGAAAACAATGCTTCA 6152

Qy     262 --------------------AlaGlyTyrValAspThrProTrpTyrLysLysTrpTyr 274
                                 |||::::::::::         ||||||::::::
Db    6153 TTAGTTGAAAAAGCTGGGCGAGAAGGATTCATTGAAAATAAGCCATATAAACAGTTTAAA 6212

Qy     275 MetLysLeuLysAsnPheMetVal----------------------------------- 282
           |||:::||||||   :::
Db    6213 GAAATGCTTGAAAATTTCTTCATCGAAATCGCAAGAGATTTCTTTAAGGACGATGGCGAT 6272

Qy     283 ---AsnArgValPheIleProThrLysLysPhePheAsnLysGlu---------IleArg 298
              :::   :::|||:::   ||||||:::      |||:::|||         :::
Db    6273 ATGTCTGAATTATTTGTTGAGACAAAGCAACGTAGAAATGAAGAACATGATTTGTTATCT 6332

Qy     299 GluProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGluAsn 318
           :::   ||||||      |||   |||               |||||||||:::
Db    6333 AAAAGATCTAAACAAACTAAAGCTAAAAAAGATAGATTAAAGAAAGATCTGTAT------ 6386

Qy     319 LysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspProSerLysAla 338
                                 |||||||||:::|||         |||
Db    6387 --------------------GATTTTTTTGATAAGTTAGATAATGATTACTGGAATATT 6425

Qy     339 LeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsn---LysIleGlyGln 357
                 |||:::   |||      ::::::   |||   :::   :::|||
Db    6426 GAAATAAATAAGCTAATCAATAAAAACGAGGAATATTTCTCCAGTACAGAAATAACAGAC 6485

Qy     358 GlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLys 377
                ::::|||::::::   |||:::|||::::::   :::   |||:::|||:::
Db    6486 ACCAATATAGATTATGTATACAATAAAATTAAAGAACAAAATGATGCTATCATTAAAAAT 6545

Qy     378 ValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhe 397
           :::            :::         :::          |||   |||
Db    6546 CTACGTAATTCTGTGGAT------ATAAAGAAACCCTCTGGAGTTGGATTAACAAAAGAG 6599

Qy     398 IleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValTyrThrGluAla 417
           :::::::|||   :::   |||      :::          :::|||:::
Db    6600 TTATCTAAT---TTATGGGATAGATATCAAATAGAAAGACAAAAAATACTGTTATCACTA 6656

Qy     418 AspAspLeuPheGluAsn------------------LysIleGlyGlnGlyThrValAsp 431
           :::::::|||   :::|||                 ::::::                |||
Db    6657 AATGAGCTAAAAGATAACGTTGATAGAAAGCTTATAGAACTGGATAATAAAAATAATGAT 6716

Qy     432 PheIleAsn------LysGluIleArgAsp----------------------------- 439
           |||:::|||      |||   :::   |||
Db    6717 TTTCTCAACTTACGGAAGAGACTTGAAGATTCTTTGAATCTACAACAAAGTTACTATGAA 6776

Qy     440 ------------------------ProSerLysAlaLeuIleArgLysValSerThrGlu 451
                                    :::|||      ::::::   |||   ::::::
Db    6777 AAAGAACTAACAAAGTTATATAATGACGCTAAAAATGCTTTGAAAGATGTGCAATCTAAA 6836

Qy     452 AlaAspAsnLeuLeu 456
           |||:::   |||:::
Db    6837 GCAAATAGGTTAATT 6851

RESULT 15
AAV71915
ID   AAV71915 standard; DNA; 3883 BP.
XX
AC   AAV71915;
XX
DT   11-FEB-1999  (first entry)
XX
DE   S. cerevisiae CIN8 DNA sequence.
XX
KW   TIH1; recombinant; research; epitope mapping; modulating; CKI; yeast;
KW   casein kinase I; cell growth; CIN8; cancer; viral infection; ss.
XX
OS   Saccharomyces cerevisiae.
XX
PN   US5846764-A.
```

```
XX
PD   08-DEC-1998.
XX
PF   23-JAN-1995;   95US-0376843.
XX
PR   23-JAN-1995;   95US-0376843.
PR   21-JAN-1994;   94US-0184605.
XX
PA   (ICOS-) ICOS CORP.
XX
PI   Demaggio AJ,  Hoekstra MF;
XX
DR   WPI; 1999-059057/05.
XX
PT   Polynucleotide encoding yeast TIH1 polypeptide - useful for
PT   producing recombinant polypeptide
XX
PS   Example 3; Columns 239-244 (columns 65-70); 46pp; English.
XX
CC   This represents the DNA sequence of Saccharomyces cerevisiae CIN8 that
CC   belongs to the yeast kinesin gene family. The invention relates to TIH
CC   proteins that interact with casein kinase I (CKI) isoforms. A host cell
CC   transformed with a DNA expression construct containing the TIH1 DNA can
CC   be used to produce recombinant TIH1 polypeptide which is used for
CC   research purposes. The TIH proteins are used to map TIH epitopes in the
CC   development of epitope-specific agents that may be useful for modulating
CC   CKI/TIH protein interactions involved in cell growth in health and
CC   disease, e.g. cancer or viral infections.
CC   Note: The specification has erroneous column number indications.
XX
SQ   Sequence 3883 BP; 1495 A; 652 C; 680 G; 1056 T; 0 other;

Alignment Scores:
Pred. No.:                0.115          Length:          3883
Score:                    120.00         Matches:         121
Percent Similarity:       34.75%         Conservative:    92
Best Local Similarity:    19.74%         Mismatches:      214
Query Match:              5.09%          Indels:          186
DB:                       20             Gaps:            29

US-09-807-459-2 (1-458) x AAV71915 (1-3883)

Qy       4 SerAspSerValGlyAspValThrLysThrLeuLeuAlaAlaSerGluSerValAspSer 23
Db    1051 AGCGATGCAGCAGGAATTATACCGAGGGTTCTTTTGAAGTTGTTTGACACATTGGAACTA 1110

Qy      24 AlaAlaAsnAlaTyrMetIle----------AsnSerAspMetSer 35
Db    1111 CAACAGAACGATTACGTAGTAAAATGTTCGTTCATTGAACTCTACAACGAAGAATTGAAG 1170

Qy      36 AspTyrLeuSerAlaValSerAsp----------AsnPheAla 46
Db    1171 GACCTCTTGGACAGCAATAGCAACGGCTCTAGTAATACTGGCTTTGACGGCCAATTTATG 1230

Qy      47 GluArgIleCysSerGlnValProLysGlySerAsnCysSerAlaSerValSerAlaTyr 66
Db    1231 AAAAAATTGAGGATTTTTGCTTCAAGCACAGCAAATAATACCACTAGCAACAGTGCTAGT 1290

Qy      67 MetSerArgCysAlaLysGlnAspCys----------LeuThrLeuGln 79
Db    1291 AGTTCCAGGAGTAATTCTAGGAACAGTTCTCCGAGGTCATTAAATGATCTAACACCTAAA 1350

Qy      80 Ser------LeuLysTyrProLeuGluAlaLysTyrGlnProLeuThrLeuProAspPro 97
Db    1351 GCTGCTCTATTAAGAAAAAGGTTAAGGACAAAATCACTGCCGAATACCATCAAGCAACAG 1410

Qy      98 TyrGlnLeuGluAlaAlaPheIleLeuPheLysGluSerAspAlaAsnProAlaAsnSer 117
Db    1411 TATCAACAACAACAGGCAGTGAATTCCAGGAACAACTCTTCCTCTAACTCTGGCTCTACC 1470

Qy     118 ThrGlu----------LysArgPheTrpMetArgPhe 126
Db    1471 ACTAATAATGCTTCTAGTAACACCAACACAAATAACGGTCAAAGAAGTTCGATGGCTCCA 1530

Qy     127 ArgArgGlyLysAsnHisSerTyrPheHisAspLeuVal---------- 139
Db    1531 AATGACCAAACTAATGGTATATACATCCAGAATTTGCAAGAATTTCACATAACAAATGCT 1590

Qy     140 ---------PheAsnLeuLeuGluLysAsnValThr---ArgAspAlaAspAlaThrAsp 155
Db    1591 ATGGAGGGGCTAAACCTATTACAAAAAGGCTTAAAGCATAGGCAAGTAGCGTCCACTAAA 1650

Qy     156 IleGluAsnPheAlaSerArg------TyrLeuTyrMetAlaThrLeuTyrTyrLysThr 173
Db    1651 ATGAACGATTTTTCCAGTAGATCTCATACCATTTTTACAATCACTTTGTATAAGAAGCAT 1710

Qy     174 Tyr----------ThrAsn 176
Db    1711 CAGGATGAACTATTTAGAATTTCCAAAATGAATCTTGTGGATTTAGCTGGTTCAGAAAAC 1770

Qy     177 ValAspGluPheGly----------AlaSerPhePheAsnLysLeu 188
Db    1771 ATCAACAGATCCGGAGCATTAAATCAACGTGCCAAAGAAGCTGGTTCAATCAACCAAAGT 1830

Qy     189 SerPheThrThrGly---------- 193
Db    1831 CTATTGACGCTGGGCAGGGTCATAAACGCACTCGTAGATAAAAGCGGCCATATACCTTTC 1890

Qy     194 ----------LeuPheGlyTrpGlyIleLysArgAla 202
Db    1891 CGTGAATCGAAATTGACCCGCCTGCTTCAAGATTCCCTGGGTGGTAATACGAAAACCGCA 1950

Qy     203 LeuLysGlnIleIleArgSerAsnLeuProLeuAspIleGlyThrGluHisSerValSer 222
Db    1951 CTA---------ATTGCTACTATATCGCCTGCAAAGGTAACTTCTGAAGAAACCTGCAGT 2001

Qy     223 ArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaLeu 242
Db    2002 ACATTAGAGTAT---GCTTCGAAGGCTAAAAACATTAAGAACAAG------CCGCAACTG 2052

Qy     243 ProLysPheAlaLysArgPheSerLeuMet----------Val 253
Db    2053 GGTTCATTTATAATGAAGGATATTTTGGTTAAAAATATAACTATGGAATTAGCAAAGATT 2112

Qy     254 ValGlnArgLeuLeuAlaThrValAlaGly---------TyrValAspThrProTrpTyr 270
Db    2113 AAATCCGATTTACTCTCTACAAAGTCCAAAGAAGGAATATATATGAGCCAAGATCACTAC 2172

Qy     271 LysLysTrpTyrMetLysLeuLysAsnPheMetValAsnArgValPheIleProThrLys 290
Db    2173 AAAAATTTGAACAGTGATTTAGAAAGTTATAAA---AATGAAGTT---------- 2214

Qy     291 LysPhePheAsnLysGluIleArgGluProSerLysAlaLeuLysGluLysValSerThr 310
Db    2215 ----------CAAGAATGTAAAAGAGAAATTGAAAGTTTGACATCGAAAAATGCATTG 2262

Qy     311 AspThrLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLys 330
Db    2263 CTAGTAAAAGATAAATTGAAGTCAAAA---------GAAACTATTCAATCTCAAAATTGC 2313

Qy     331 GluIleArgAspProSerLysAlaLeu----------LysGlu 341
Db    2314 CAAATAGAATCATTGAAAACTACCATAGATCATTTAAGGGCACAACTAGATAAACAGCAT 2373

Qy     342 LysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThr------ 359
Db    2374 AAAACTGAAATTGAAATATCCGAT---TTTAATAACAAACTACAGAAGTTGACTGAGGTA 2430

Qy     360 ----------ValAspPhe---------- 362
Db    2431 ATGCAAATGGCCCTACATGATTACAAAAAAAGAGAACTTGACCTTAATCAAAAGTTTGAA 2490

Qy     363 ------IleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValSerThr 380
Db    2491 ATGCATATTACTAAAGAAATTAAAAAATTGAAATCTACACTGTTTTTACAATTAAACACT 2550
```

```
Qy     381 GlyAlaGluAsp------LeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIle 398
Db    2551 ATGCAACAGGAAAGTATTCTTCAAGAGACTAATATC---CAACCAAATCTTGATATGATC 2607

Qy     399 AsnAsnGluIleArgAspProSerLysAlaLeu-------------------------- 409
Db    2608 AAAAATGAAGTACTGACTCTTATGAGAACCATGCAAGAAAAAGCTGAACTAATGTACAAA 2667

Qy     410 ------IleArgLysValTyrThrGluAlaAspAspLeuPheGlu---------AsnLys 424
Db    2668 GACTGTGTGAAGAAAATTTTAAACGAATCTCCTAAATTCTTCAATGTTGTTATTGAGAAA 2727

Qy     425 IleGlyGlnGlyThrValAspPhe------IleAsnLysGluIleArgAspProSerLys 442
Db    2728 ATCGACATAATAAGAGTAGATTTCCAAAAATTTTATAAAAATATAGCCGAGAAT------ 2781

Qy     443 AlaLeuIleArgLysValSerThrGluAlaAspAsnLeu 455
Db    2782 ------CTTTCTGATATTAGCGAAGAAAATAACAACATG 2814
```

```
RESULT 16
AAH78010
ID   AAH78010 standard; DNA; 3884 BP.
XX
AC   AAH78010;
XX
DT   13-NOV-2001  (first entry)
XX
DE   Nucleotide sequence of bimC homologue, cin8.
XX
KW   bimC; kinesin related protein; fungal viability; antifungal; cin8;
KW   fungal infection; ss.
XX
OS   Saccharomyces cerevisiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             491..3607
FT                   /*tag= a
FT                   /product= "cin8"
XX
PN   US6284480-B1.
XX
PD   04-SEP-2001.
XX
PF   03-APR-2000; 2000US-0541782.
XX
PR   03-APR-2000; 2000US-0541782.
XX
PA   (CYTO-) CYTOKINETICS INC.
XX
PI   Nislow CE,  Sakowicz R,  Beraud C;
XX
DR   WPI; 2001-540724/60.
DR   P-PSDB; AAG67416.
XX
PT   Identifying a modulator, e.g. antifungal agent, of a target protein
PT   comprising bimC or its fragment by determining enzymatic activity of a
PT   reaction, in the presence and absence of the compound, that uses ADP or
PT   phosphate produced by bimC -
XX
PS   Disclosure; Fig 3A-B; 47pp; English.
XX
CC   The present sequence encodes a bimC homologue, designated cin8. BimC is
CC   a kinesin related protein, which is essential for fungal viability. The
CC   specification describes a method of identifying modulators of bimC.
CC   The method comprises adding a test agent to a mixture comprising bimC
CC   protein that directly or indirectly produces ADP or phosphate, subjecting
CC   the mixture to an enzymatic reaction that uses the ADP or phosphate,
CC   and determining the enzymatic activity in presence and absence of test
CC   compound. A change in the activity level between the presence and absence
CC   of the candidate agent indicates a modulator of the target protein
CC   function. The method is useful for identifying a modulator, e.g.
CC   antifungal agents, of bimC. The modulators can be used, for example, to
CC   inhibit the growth or spread of fungi, mould, fruit flies, etc.. The
CC   modulators can be used for preventing and treating infections caused
CC   by Chytridiomycetes, Hyphochrytridiomycetes, Plasmodiophoromycetes,
CC   Oomycetes, Zygomycetes, Ascomycetes, and Basidiomycetes.
XX
SQ   Sequence 3884 BP; 1496 A; 652 C; 680 G; 1056 T; 0 other;
```

```
Alignment Scores:
Pred. No.:                0.115          Length:          3884
Score:                    120.00         Matches:         121
Percent Similarity:       34.75%         Conservative:    92
Best Local Similarity:    19.74%         Mismatches:      214
Query Match:              5.09%          Indels:          186
DB:                       22             Gaps:            29

US-09-807-459-2 (1-458) x AAH78010 (1-3884)

Qy       4 SerAspSerValGlyAspValThrLysThrLeuLeuAlaAlaSerGluSerValAspSer 23
Db    1052 AGCGATGCAGCAGGAATTATACCGAGGGTTCTTTTGAAGTTGTTTGACACATTGGAACTA 1111

Qy      24 AlaAlaAsnAlaTyrMetIle----------------------AsnSerAspMetSer 35
Db    1112 CAACAGAACGATTACGTAGTAAAATGTTCGTTCATTGAACTCTACAACGAAGAATTGAAG 1171

Qy      36 AspTyrLeuSerAlaValSerAsp--------------------------AsnPheAla 46
Db    1172 GACCTCTTGGACAGCAATAGCAACGGCTCTAGTAATACTGGCTTTGACGGCCAATTTATG 1231

Qy      47 GluArgIleCysSerGlnValProLysGlySerAsnCysSerAlaSerValSerAlaTyr 66
Db    1232 AAAAAATTGAGGATTTTTGCTTCAAGCACAGCAAATAATACCACTAGCAACAGTGCTAGT 1291

Qy      67 MetSerArgCysAlaLysGlnAspCys--------------------LeuThrLeuGln 79
Db    1292 AGTTCCAGGAGTAATTCTAGGAACAGTTCTCCGAGGTCATTAAATGATCTAACACCTAAA 1351

Qy      80 Ser------LeuLysTyrProLeuGluAlaLysTyrGlnProLeuThrLeuProAspPro 97
Db    1352 GCTGCTCTATTAAGAAAAAGGTTAAGGACAAAATCACTGCCGAATACCATCAAGCAACAG 1411

Qy      98 TyrGlnLeuGluAlaAlaPheIleLeuPheLysGluSerAspAlaAsnProAlaAsnSer 117
Db    1412 TATCAACAACAACAGGCAGTGAATTCCAGGAACAACTCTTCCTCTAACTCTGGCTCTACC 1471

Qy     118 ThrGlu---------------------------------LysArgPheTrpMetArgPhe 126
Db    1472 ACTAATAATGCTTCTAGTAACACCAACACAAATAACGGTCAAAGAAGTTCGATGGCTCCA 1531

Qy     127 ArgArgGlyLysAsnHisSerTyrPheHisAspLeuVal--------------------- 139
Db    1532 AATGACCAAACTAATGGTATATACATCCAGAATTTGCAAGAATTTCACATAACAAATGCT 1591

Qy     140 ---------PheAsnLeuLeuGluLysAsnValThr---ArgAspAlaAspAlaThrAsp 155
Db    1592 ATGGAGGGGCTAAACCTATTACAAAAAGGCTTAAAGCATAGGCAAGTAGCGTCCACTAAA 1651

Qy     156 IleGluAsnPheAlaSerArg------TyrLeuTyrMetAlaThrLeuTyrTyrLysThr 173
Db    1652 ATGAACGATTTTTCCAGTAGATCTCATACCATTTTTACAATCACTTTGTATAAGAAGCAT 1711

Qy     174 Tyr-------------------------------------------------------ThrAsn 176
Db    1712 CAGGATGAACTATTTAGAATTTCCAAAATGAATCTTGTGGATTTAGCTGGTTCAGAAAAC 1771

Qy     177 ValAspGluPheGly-----------------------AlaSerPhePheAsnLysLeu 188
Db    1772 ATCAACAGATCCGGAGCATTAAATCAACGTGCCAAAGAAGCTGGTTCAATCAACCAAAGT 1831

Qy     189 SerPheThrThrGly------------------------------------------- 193
Db    1832 CTATTGACGCTGGGCAGGGTCATAAACGCACTCGTAGATAAAAGCGGCCATATACCTTTC 1891
```

```
Qy     194 ------------------------------LeuPheGlyTrpGlyIleLysArgAla 202
                                         |||         |||   |||
Db    1892 CGTGAATCGAAATTGACCCGCCTGCTTCAAGATTCCCTGGGTGGTAATACGAAAACCGCA 1951

Qy     203 LeuLysGlnIleIleArgSerAsnLeuProLeuAspIleGlyThrGluHisSerValSer 222
           |||         |||   :::      |||      :::   :::|||   :::   |||
Db    1952 CTA---------ATTGCTACTATATCGCCTGCAAAGGTAACTTCTGAAGAAACCTGCAGT 2002

Qy     223 ArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaLeu 242
              |||::::::      |||      |||:::      :::         |||   |||
Db    2003 ACATTAGAGTAT---GCTTCGAAGGCTAAAAACATTAAGAACAAG------CCGCAACTG 2053

Qy     243 ProLysPheAlaLysArgPheSerLeuMet--------------------------Val 253
           |||      :::      |||:::                              :::
Db    2054 GGTTCATTTATAATGAAGGATATTTTGGTTAAAAATATAACTATGGAATTAGCAAAGATT 2113

Qy     254 ValGlnArgLeuLeuAlaThrValAlaGly---------TyrValAspThrProTrpTyr 270
                 ||||||:::|||   :::            |||:::            |||
Db    2114 AAATCCGATTTACTCTCTACAAAGTCCAAAGAAGGAATATATATGAGCCAAGATCACTAC 2173

Qy     271 LysLysTrpTyrMetLysLeuLysAsnPheMetValAsnArgValPheIleProThrLys 290
           |||          |||::::::::::      |||   |||
Db    2174 AAAAATTTGAACAGTGATTTAGAAAGTTATAAA---AATGAAGTT--------------- 2215

Qy     291 LysPheAsnLysGluIleArgGluProSerLysAlaLeuLysGluLysValSerThr 310
                 :::|||   :::         ::::::|||      |||   :::
Db    2216 -----------CAAGAATGTAAAAGAGAAATTGAAAGTTTGACATCGAAAAATGCATTG 2263

Qy     311 AspThrLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLys 330
                |||||      ::::::|||            |||:::         |||
Db    2264 CTAGTAAAAGATAAATTGAAGTCAAAA---------GAAACTATTCAATCTCAAAATTGC 2314

Qy     331 GluIleArgAspProSerLysAlaLeu--------------------------LysGlu 341
           :::|||                  :::                          :::
Db    2315 CAAATAGAATCATTGAAAACTACCATAGATCATTTAAGGGCACAACTAGATAAACAGCAT 2374

Qy     342 LysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThr------ 359
           |||      :::    |||   |||   ||||||:::   :::   |||
Db    2375 AAAACTGAAATTGAAATATCCGAT---TTTAATAACAAACTACAGAAGTTGACTGAGGTA 2431

Qy     360 ----------------------------------ValAspPhe--------------- 362
                                               :::|||
Db    2432 ATGCAAATGGCCCTACATGATTACAAAAAAAGAGAACTTGACCTTAATCAAAAGTTTGAA 2491

Qy     363 ------IleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValSerThr 380
                |||      ||||||:::               |||      ::::::::::|||
Db    2492 ATGCATATTACTAAAGAAATTAAAAAATTGAAATCTACACTGTTTTTACAATTAAACACT 2551

Qy     381 GlyAlaGluAsp------LeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIle 398
           ::::::      |||   |||      |||   |||      :::|||   |||
Db    2552 ATGCAACAGGAAAGTATTCTTCAAGAGACTAATATC---CAACCAAATCTTGATATGATC 2608

Qy     399 AsnAsnGluIleArgAspProSerLysAlaLeu--------------------------- 409
             ||||||:::          :::   :::
Db    2609 AAAAATGAAGTACTGACTCTTATGAGAACCATGCAAGAAAAAGCTGAACTAATGTACAAA 2668

Qy     410 ------IleArgLysValTyrThrGluAlaAspAspLeuPheGlu---------AsnLys 424
                ::::::|||:::      |||:::         |||               |||
Db    2669 GACTGTGTGAAGAAAATTTTAAACGAATCTCCTAAATTCTTCAATGTTGTTATTGAGAAA 2728

Qy     425 IleGlyGlnGlyThrValAspPhe------IleAsnLysGluIleArgAspProSerLys 442
           |||            |||||||||           |||   |||   :::
Db    2729 ATCGACATAATAAGAGTAGATTTCCAAAAATTTTATAAAAATATAGCCGAGAAT------ 2782

Qy     443 AlaLeuIleArgLysValSerThrGluAlaAspAsnLeu 455
                :::      :::|||   |||   :::|||:::
Db    2783 ------CTTTCTGATATTAGCGAAGAAAATAACAACATG 2815

RESULT 17
AAQ70102
ID   AAQ70102 standard; cDNA to mRNA; 4766 BP.
XX
AC   AAQ70102;
XX
DT   22-SEP-1994  (first entry)
XX
DE   Malarial PfEMP3 epitopic fragment clone p2b1;p12-1.
XX
KW   Plasmodium falciparum erythrocyte membrane protein; PfEMP3;
KW   malaria; antigen; epitope; vaccine; anti-idiotype antibody; ds.
XX
OS   Plasmodium falciparum (Malayan Camp strain).
XX
FH   Key             Location/Qualifiers
FT   CDS             3..4766
FT                   /*tag= a
FT                   /note= "partial coding region; does not include
FT                        initiation or termination codons"
XX
PN   WO9403604-A.
XX
PD   17-FEB-1994.
XX
PF   05-AUG-1993;   93WO-US07261.
XX
PR   07-AUG-1992;   92US-0927531.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Handunnetti SM,  Howard RJ,  Pasloske BL,  Van Schravendijk MR;
XX
DR   WPI; 1994-065693/08.
DR   P-PSDB; AAR46605.
XX
PT   New malaria antigen, PfEMP3 - used to isolate and produce prods.
PT   for use in diagnosis, therapy and prevention of malarial
PT   infection
XX
PS   Claim 1; Page 72-79; 79pp; English.
XX
CC   The PfEMP3 malarial antigen is recognised by monoclonal antibody MAb
CC   12C11. Nucleic acid sequences encoding part of the 315kD antigen,
CC   have been isolated and sequenced. PfEMP3 is encoded on chromosome 2
CC   of the P.falciparum genome and is thought to be associated with knob
CC   formation and structure; malarial strains carrying deletions of the
CC   gene coding for PfEMP3 exhibit a knobless phenotype.
XX
SQ   Sequence 4766 BP; 2404 A; 508 C; 863 G; 991 T; 0 other;

Alignment Scores:
Pred. No.:                0.212          Length:          4766
Score:                    118.50         Matches:         57
Percent Similarity:       47.57%         Conservative:    31
Best Local Similarity:    30.81%         Mismatches:      66
Query Match:              5.02%          Indels:          31
DB:                       15             Gaps:            14

US-09-807-459-2 (1-458) x AAQ70102 (1-4766)

Qy     294 AsnLysGluIleArg---GluProSerLysAlaLeuLysGluLysValSerThrAspThr 312
           |||||||||:::|||   :::   |||:::   |||||||||
Db    3135 AATAAAGAATTACGAAATAAAGGATCTGAAGGATTAAAAGAAAATGCAGAACTAAAAAAT 3194

Qy     313 LysAspLeuPheGluAsnLysIleGlyGlnGly---------ThrValAspPhePheAsn 329
           |||:::|||   :::||||||      :::|||                :::      |||
Db    3195 AAAGAATTA---CAAAATAAAGGATCTGAAGGATTAAAAGAAAATGCAGAATTAAAAAAT 3251

Qy     330 LysGluIleArgAspPro---SerLysAlaLeuLysGluLysValSerAsnAspAlaLys 348
           ||||||:::::::::      |||:::   |||||||||                  |||
Db    3252 AAAGAATTACAAAATAAAGGATCTGAAGGATTAAAAGAAAATGCAGAGCAAAAAAATAAA 3311

Qy     349 AspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArg 368
           :::|||   :::||||||      :::|||      :::      |||   |||::::::
Db    3312 GAATTA---CAAAATAAAGGATCTGAAGGATTAAAAGAA------AATGCAGAATTAAAA 3362

Qy     369 AspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGlu--- 387
```

```
Db   3363 ------AATAAAGAATTACGAAATAAAGGATCTGATGGATTAAAAGAAAATGCAGAGCTA 3416

Qy    388 -----------------AsnLysIleGlyGlnGly---------ThrValAspPheIle 398
Db   3417 AAAAATAAAGAATTACGAAATAAAGGATCTGATGGATTAAAAGAAAATGCAGAATTAAAA 3476

Qy    399 AsnAsnGluIleArgAspPro---SerLysAlaLeuIleArgLysValTyrThrGluAla 417
Db   3477 AATAAAGAATTACGAAATAAAGGATCTGAAGGATTAAAAGAAAATGTATATACAAATAAT 3536

Qy    418 AspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnLysGluIle 437
Db   3537 GAT------TTAAAGAATAACGATAATCAAAATAAA---GATTTATCTAATAAAGACATG 3587

Qy    438 ArgAspProSerLysAlaLeuIleArgLysValSerThrGluAlaAsp------------ 453
Db   3588 AAA------AATAAAGAACTATTAAACAAAGATATTTCTAATAAAGACATGAAGAATAAA 3641

Qy    454 AsnLeuLeuGluLys 458
Db   3642 GAACTATTAAATAAA 3656

RESULT 18
AAA70095
ID   AAA70095 standard; DNA; 7326 BP.
XX
AC   AAA70095;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:228.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 454-456; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence 7326 BP; 3588 A; 868 C; 1321 G; 1549 T; 0 other;

Alignment Scores:
Pred. No.:              0.377          Length:          7326
Score:                  118.50         Matches:         58
Percent Similarity:     48.65%         Conservative:    32
Best Local Similarity:  31.35%         Mismatches:      64
Query Match:            5.02%          Indels:          31
DB:                     21             Gaps:            14

US-09-807-459-2 (1-458) x AAA70095 (1-7326)

Qy    294 AsnLysGluIleArg---GluProSerLysAlaLeuLysGluLysValSerThrAspThr 312
Db   3361 AATAAAGAGTTACAAAATAAAGGATCGGAAGGATTAAAAGAAAATGCAGAATTAAAAAAT 3420

Qy    313 LysAspLeuPheGluAsnLysIleGlyGlnGly---------ThrValAspPhePheAsn 329
Db   3421 AAAGAGTTA---CAAAATAAAGGATCGGAAGGATTAAAAGAAAATGCAGAATTAAAAAAT 3477

Qy    330 LysGluIleArgAspPro---SerLysAlaLeuLysGluLysValSerAsnAspAlaLys 348
Db   3478 AAAGAATTACGAAATAAAGGATCTGAAGGATTAAAAGAAAATGCAGAGCTAAAAAATAAA 3537

Qy    349 AspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArg 368
Db   3538 GAATTA---CAAAATAAAGGATCTGAAGGATTAAAAGAA------AATGCAGAATTAAAA 3588

Qy    369 AspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGlu--- 387
Db   3589 ------AATAAAGAGTTACAAAATAAAGGATCGGAAGGATTAAAAGAAAATGCAGAATTA 3642

Qy    388 -----------------AsnLysIleGlyGlnGly---------ThrValAspPheIle 398
Db   3643 AAAAATAAAGAGTTACAAAATAAAGGATCGGAAGGATTAAAAGAAAATGCAGAATTAAAA 3702

Qy    399 AsnAsnGluIleArgAspPro---SerLysAlaLeuIleArgLysValTyrThrGluAla 417
Db   3703 AATAAAGAATTACGAAATAAAGGATCTGAAGGATTAAAAGAAAATGTATATACAAAT--- 3759

Qy    418 AspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnLysGluIle 437
Db   3760 AATGATTTAAAGAATAACGATATT------CAAAATAAAGATTTATCTAATAAAGACATG 3813

Qy    438 ArgAspProSerLysAlaLeuIleArgLysValSerThrGluAlaAsp------------ 453
Db   3814 AAA------AATAAAGAACTATTAAACAAAGATATTTCTAATAAAGACATGAAGAATAAA 3867

Qy    454 AsnLeuLeuGluLys 458
Db   3868 GAACTATTAAATAAA 3882

RESULT 19
AAA81514/c
ID   AAA81514 standard; DNA; 33303 BP.
XX
AC   AAA81514;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_61 SEQ ID NO:61.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
```

```
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 1375-1385; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 33303 BP; 7919 A; 8691 C; 8422 G; 8266 T; 5 other;

Alignment Scores:
Pred. No.:              4           Length:         33303
Score:                  117.00      Matches:        81
Percent Similarity:     35.66%      Conservative:   62
Best Local Similarity:  20.20%      Mismatches:     150
Query Match:            4.96%       Indels:         108
DB:                     21          Gaps:           19

US-09-807-459-2 (1-458) x AAA81514 (1-33303)

Qy      94 LeuProAspProTyrGlnLeuGluAlaAlaPheIleLeuPheLysGluSerAspAlaAsn 113
           |||   |||   |||      :::||||||   |||:::   |||::::::::::   |||
Db  33123 TTGGGTGACCGTTATGCGGATGATGCTGCTGCAATTGTCGGTAAGGATGCAAACTTAAAT 33064

Qy     114 ProAlaAsnSerThrGluLysArgPheTrpMetArgPheArgArgGlyLysAsnHisSer 133
               |||                 ||||||      :::::::|||
Db  33063 GGTTTGAAT--------------TTATGGATG------AAAAAAGGT------------ 33037

Qy     134 TyrPheHisAspLeuValPheAsnLeuLeuGluLysAsnValThrArgAspAlaAspAla 153
                       |||   ||||||   :::      |||   :::
Db  33036 --------------GTGGAAAACCTATGGGATGATACGGTCGGTAAA---------AAG 33001

Qy     154 ThrAspIleGluAsnPheAlaSerArgTyrLeuTyrMetAlaThrLeuTyrTyrLysThr 173
           |||   :::|||   |||                :::|||      :::::::::
Db  33000 ACCCGTTTAGAGAAATTTGATCGG-----------GTTGCACTGCAACATTTCAGGCAA 32953

Qy     174 TyrThr--------------------------------AsnValAspGluPheGlyAla 182
           |||                                   |||      |||
Db  32952 TATGCGCGTCTAATTAATCAAAATAATGGTAGATTACCCAATACTAGTGAAATTGAGAGA 32893

Qy     183 SerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAla 202
           |||:::::      ::::::                  |||                |||
Db  32892 AGTTACTATAAAGCCGTTACCGATAAT-----------GGCGTTTCTTCCAGTGCAGCT 32845

Qy     203 LeuLysGlnIleIleArgSerAsnLeuPro----------------------LeuAsp 214
           :::      :::|||      :::||||||                      |||
Db  32844 ATTGATTTAGTTATTAATCGTTCACTTCCGGATATGGCGGATGGTTATTGGGCATTAGGT 32785

Qy     215 IleGlyThrGlu-----------HisSerValSerArgLeuGlnHisIleThrSerSer 230
           :::|||   |||           |||:::      :::   :::         |||
Db  32784 TTGGGGATAGAAGCCGAACGTATCCACAATGAGCAAGCAGTAAATAATCCGAACGGTAGC 32725

Qy     231 TyrLysAspTyrMetAspThrGlnIleProAlaLeuProLys----------------- 244
              :::|||               |||   ||||||   |||
Db  32724 GAAAGGGATAATAGAAAGCAGTTAATATCTGCTTTAGATAAAGGATTTGATGGATCTTTT 32665

Qy     245 PheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValAlaGly--- 263
                 |||   |||:::   :::            :::      |||      |||
Db  32664 AAAGAGAAGCATTTTACTTTTTTACAATCTGTGATGATGGATGTAACAAAGTTAGGTGTT 32605

Qy     264 ---TyrValAspThrProTrpTyrLys------LysTrp-----------TyrMetLys 276
              |||           |||   |||         |||
Db  32604 GAATATACAATAGATGGTTGGCAAAAAATTGGAGGTTGGGGTAATGGGATAATCAATGAT 32545

Qy     277 LeuLysAsnPheMetValAsnArgValPheIleProThrLysLysPhePheAsnLysGlu 296
           |||          ::|||   |||   :::            :::      |||
Db  32544 TTATATAAAAGTGTTGTAAAAAGAGAGTGGACTGCAATATTTGAGATCGTTAATAATAAC 32485

Qy     297 IleArgGluProSerLysAlaLeuLysGluLysValSer---------------------- 309
           |||::::::   :::::|||   |||   :::::::::
Db  32484 ATCAAGCAAGGAAATGAAGCTTTTAAAAATGAAATCAATAGCTTGGTTCATGATATGAAA 32425

Qy     310 ------------------------------------------------ThrAspThr 312
                                                            |||
Db  32424 GCTGCTGGCAAGGAATTTGGAGATGACTTAAATACACAGTGGAATAATCTCACTCAGGCT 32365

Qy     313 LysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIle 332
              ::::::      |||   |||         |||            |||   :::
Db  32364 GCCGAAATAATCTATAATGACATAGTAGACAATACTAGTCAAGGAATAGAAAAAGGTGTC 32305

Qy     333 ArgAspProSerLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGlu 352
           :::         |||   |||   |||||::   |||   |||   ||||||   :::
Db  32304 AAA---GCCATTAAAGAATTGTCTGAAAAAATGAAAAATGCTGCTTCCGATTTGGCTGAC 32248

Qy     353 AsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLys 372
                                                          :::   :::|||
Db  32247 GGTTCA-----------------------------------------GCAGAGAAAGCTAAA 32227

Qy     373 AlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGln 392
              ::::::      ::::::      |||::::::   :::||||||      :::
Db  32226 CAAGTAGTGGAAGATTTGGCTCAAGCCGCCAAAGAAGCATACGAAAAT------GCCAAA 32173

Qy     393 GlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeu------Ile 410
           |||   :::                |||:::      |||   |||
Db  32172 TCCACAGCCGAGAAGGCTGCTCAAGCAGCTCGAGAATTTTTTAAGGGCTTGCCCAGTTTT 32113

Qy     411 ArgLysValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrVal 430
           :::   :::      :::      |||||||||   |||         :::|||   :::
Db  32112 AAAGATCTGGCCGAAAAATTTAGAGATCTGTTCCCAAATCCG------GAAGGCTGGATC 32059
```

```
Qy    431 Asp 431
          |||
Db  32058 GAT 32056

RESULT 20
AAF21610
ID   AAF21610 standard; DNA; 349980 BP.
XX
AC   AAF21610;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Neisseria meningitidis B nucleotide sequence SEQ ID NO:111.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; immunogenic; vaccine;
KW   diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW   ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200066791-A1.
XX
PD   09-NOV-2000.
XX
PF   08-MAR-2000; 2000WO-US05928.
XX
PR   30-APR-1999;   99US-0132068.
PR   08-OCT-1999;   99WO-US23573.
PR   28-FEB-2000; 2000GB-0004695.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Pizza M,  Hickey E,  Peterson J,  Tettelin H,  Venter JC,  Masignani V;
PI   Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V,  Rappuoli R;
PI   Frazer CM,  Grandi G;
XX
DR   WPI; 2000-647603/62.
XX
PT   Neisseria meningitidis B full length genome sequence and open reading
PT   frames are used to detect, treat and prevent Neisserial infections -
XX
PS   Claim 7; Appendix A; 692pp; English.
XX
CC   The present invention describes the full length genome of
CC   Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC   to AAF21613 represent fragments of the NMB genomic sequence, as the
CC   sequence was too long to go in a record on its own it was split into 8
CC   sequences which overlap each other at the beginning and end of each
CC   sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC   the beginning  of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC   the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC   Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC   AAF21606 represent PCR primers which are used in the exemplification of
CC   the present invention. The NMB genome and fragments from it have
CC   antibacterial activity, and can be used in vaccines and gene therapy.
CC   Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC   proteins can be used in compositions for treating or preventing infection
CC   due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC   presence of Neisserial bacteria or of antibodies raised to Neisserial
CC   bacteria. Computers, computer memory, computer storage medium or computer
CC   databases can be used in a search to identify open reading frames (ORFs)
CC   or coding sequences within the NMB genome. The DNA sequences provide
CC   further opportunities to find antigenic or immunogenic proteins which are
CC   more effective in vaccines than the outer membrane proteins currently
CC   used.
XX
SQ   Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 other;

Alignment Scores:
Pred. No.:              92.3          Length:        349980
Score:                  117.00        Matches:       81
Percent Similarity:     35.66%        Conservative:  62

Best Local Similarity:  20.20%        Mismatches:    150
Query Match:            4.96%         Indels:        108
DB:                     21            Gaps:          19

US-09-807-459-2 (1-458) x AAF21610 (1-349980)

Qy      94 LeuProAspProTyrGlnLeuGluAlaAlaPheIleLeuPheLysGluSerAspAlaAsn 113
           |||   |||   |||      :::|||||||   |||:::   |||:::::::::   |||
Db 243690 TTGGGTGACCGTTATGCGGATGATGCTGCTGCAATTGTCGGTAAGGATGCAAACTTAAAT 243749

Qy     114 ProAlaAsnSerThrGluLysArgPheTrpMetArgPheArgArgGlyLysAsnHisSer 133
               |||                   ||||||      ::::::|||
Db 243750 GGTTTGAAT--------------TTATGGATG------AAAAAAGGT------------ 243776

Qy     134 TyrPheHisAspLeuValPheAsnLeuLeuGluLysAsnValThrArgAspAlaAspAla 153
                      |||   ||||||   :::      |||   :::
Db 243777 --------------GTGGAAAACCTATGGGATGATACGGTCGGTAAA---------AAG 243812

Qy     154 ThrAspIleGluAsnPheAlaSerArgTyrLeuTyrMetAlaThrLeuTyrTyrLysThr 173
           |||   :::|||   |||                 :::|||      :::::::::
Db 243813 ACCCGTTTAGAGAAATTTGATCGG-----------GTTGCACTGCAACATTTCAGGCAA 243860

Qy     174 TyrThr--------------------------------AsnValAspGluPheGlyAla 182
           |||                                   |||      |||
Db 243861 TATGCGCGTCTAATTAATCAAAATAATGGTAGATTACCCAATACTAGTGAAATTGAGAGA 243920

Qy     183 SerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAla 202
           |||::::::      ::::::                |||                |||
Db 243921 AGTTACTATAAAGCCGTTACCGATAAT-----------GGCGTTTCTTCCAGTGCAGCT 243968

Qy     203 LeuLysGlnIleIleArgSerAsnLeuPro----------------------LeuAsp 214
           :::      :::|||      :::||||||                      |||
Db 243969 ATTGATTTAGTTATTAATCGTTCACTTCCGGATATGGCGGATGGTTATTGGGCATTAGGT 244028

Qy     215 IleGlyThrGlu------------HisSerValSerArgLeuGlnHisIleThrSerSer 230
           :::|||   |||            |||:::         :::   :::         |||
Db 244029 TTGGGGATAGAAGCCGAACGTATCCACAATGAGCAAGCAGTAAATAATCCGAACGGTAGC 244088

Qy     231 TyrLysAspTyrMetAspThrGlnIleProAlaLeuProLys----------------- 244
             :::|||                |||   ||||||   |||
Db 244089 GAAAGGGATAATAGAAAGCAGTTAATATCTGCTTTAGATAAAGGATTTGATGGATCTTTT 244148

Qy     245 PheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValAlaGly--- 263
               |||   |||:::   :::             :::      |||      |||
Db 244149 AAAGAGAAGCATTTTACTTTTTTACAATCTGTGATGATGGATGTAACAAAGTTAGGTGTT 244208

Qy     264 ---TyrValAspThrProTrpTyrLys------LysTrp------------TyrMetLys 276
              |||          |||   |||         |||
Db 244209 GAATATACAATAGATGGTTGGCAAAAAATTGGAGGTTGGGGTAATGGGATAATCAATGAT 244268

Qy     277 LeuLysAsnPheMetValAsnArgValPheIleProThrLysLysPhePheAsnLysGlu 296
           |||         :::|||   |||   :::            :::      |||
Db 244269 TTATATAAAAGTGTTGTAAAAAGAGAGTGGACTGGAATATTTGAGATCGTTAATAATAAC 244328

Qy     297 IleArgGluProSerLysAlaLeuLysGluLysValSer---------------------- 309
           |||::::::   :::::::|||   |||   :::::::::
Db 244329 ATCAAGCAAGGAAATGAAGCTTTTAAAAATGAAATCAATAGCTTGGTTCATGATATGAAA 244388

Qy     310 ----------------------------------------------------ThrAspThr 312
                                                               |||
Db 244389 GCTGCTGGCAAGGAATTTGGAGATGACTTAAATACACAGTGGAATAATCTCACTCAGGCT 244448

Qy     313 LysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIle 332
             ::::::      |||   |||         |||                |||   :::
Db 244449 GCCGAAATAATCTATAATGACATAGTAGACAATACTAGTCAAGGAATAGAAAAAGGTGTC 244508

Qy     333 ArgAspProSerLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGlu 352
           :::         |||   |||   ||||||:::   |||   |||   ||||||   :::
Db 244509 AAA---GCCATTAAAGAATTGTCTGAAAAAATGAAAAATGCTGCTTCCGATTTGGCTGAC 244565

Qy     353 AsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLys 372
                                                                 :::   :::|||
Db 244566 GGTTCA-----------------------------------GCAGAGAAAGCTAAA 244586
```

```
Qy    373 AlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGln 392
Db 244587 CAAGTAGTGGAAGATTTGGCTCAAGCCGCCAAAGAAGCATACGAAAAT------GCCAAA 244640

Qy    393 GlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeu------Ile 410
Db 244641 TCCACAGCCGAGAAGGCTGCTCAAGCAGCTCGAGAATTTTTTAAGGGCTTGCCCAGTTTT 244700

Qy    411 ArgLysValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrVal 430
Db 244701 AAAGATCTGGCCGAAAAATTTAGAGATCTGTTCCCAAATCCG------GAAGGCTGGATC 244754

Qy    431 Asp 431
Db 244755 GAT 244757
```

RESULT 21
AAS71042
ID AAS71042 standard; cDNA; 6637 BP.
XX
AC AAS71042;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #6846.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG06855.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 6846; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6637 BP; 2318 A; 1415 C; 1452 G; 1452 T; 0 other;

Alignment Scores:

Pred. No.:	0.73	Length:	6637
Score:	115.00	Matches:	89
Percent Similarity:	35.76%	Conservative:	68
Best Local Similarity:	20.27%	Mismatches:	167
Query Match:	4.87%	Indels:	115
DB:	23	Gaps:	20

US-09-807-459-2 (1-458) x AAS71042 (1-6637)

```
Qy    113 AsnProAlaAsnSerThrGluLysArgPheTrpMetArgPheArgArgGlyLys-AsnHi 132
Db   5118 AATCAGGCAAGAGAAAGAAATAAAGGGTATTCAATTAGG---AAAAGAGGAAGTCAAATT 5174

Qy    132 sSerTyrPheHisAspLeuValPheAsnLeuLeuGluLysAsnValThrArgAspAlaAs 152
Db   5175 GTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGTCTCAGCCCAAAA 5234

Qy    152 pAlaThrAsp---IleGluAsnPheAlaSerArgTyrLeuTyrMetAlaThrLeuTyrTy 171
Db   5235 TCTCCTTAAGCTGATAAACAACTTCAGCAAAGTATCAGGATACAAAATCAATGTGCAAAA 5294

Qy    171 rLys---------ThrTyrThrAsnValAspGluPheGlyAlaSerPhePheAsnLysLe 188
Db   5295 ATCACAAATATTCTTATACACCAATAACAGACAAACAGAGAGCCAAATCATGAGTGAACT 5354

Qy    188 uSerPheThrThrGly--------------LeuPheGlyTrpGlyIleLysArgAlaLe 203
Db   5355 CCCATTCACAATTGCTTCAAAGACAATAAAATACCTAGGAATTCAACTTACAAGGGATGT 5414

Qy    203 uLysGlnIleIleArgSerAsnLeu---ProLeuAspIleGlyThrGluHisSerValSe 222
Db   5415 GAAGGACCTCTTCAAGGAGAATTACAAACCACTGCTCAATGAAATAAAAGAAGATACAAA 5474

Qy    222 rArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAsp---ThrGlnIleProAl 241
Db   5475 CAAATGGAACAACATTCCATGCTCATGGGTAGGAAGAATCAATATCATGAAAATGGCCAT 5534

Qy    241 aLeuProLysPheAlaLysArgPheSerLeuMetValValGln---------------- 255
Db   5535 ACTGCCCAAGGTAATTTATAGATTCAGTGCCATCGCCATCAAGCTACCAATGACTTTCTT 5594

Qy    256 ----------------------------------------------------ArgLeuLe 258
Db   5595 CACAGAACTGGAAAAAACTACTTTAAAGTTCATATGGAACCAAAAAAGAGCCCGCATTGC 5654

Qy    258 uAlaThrVal---------------------------------------------------- 261
Db   5655 CAAGTCAATCCTAAGCCAAAAGAACAAAGCCGGAGGCATCATGCTACCTGACTTCAAACT 5714

Qy    262 ----------AlaGlyTyrValAspThrProTrpTyrLysLysTrpTyrMetLys----- 276
Db   5715 ATACTACAAGGCTACAGTAACCAAAACAGCATGGTAC------TGGTACCAAAACAGAGA 5768

Qy    277 ----------------------------------------------LeuLysAsnPheMe 281
Db   5769 TATTGATCAATGGAGCAGAACAGAGCCCTCAGAAGTAATGCCACACATCTACAACTATCT 5828

Qy    281 tValAsnArgValPheIleProThrLysLysPhe--------------PheAsnLysGl 296
Db   5829 GATCTTTGATAAACTTGACAAAAACAAGAAATGGGGAAAGGATTCCCTATTTAATAAGTG 5888

Qy    296 uIleArgGluProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPh 316
Db   5889 GTGCTGGGAAAATTGGCTAGCCATATGTAGAAAGCTGAAACTGGATCCCTTCCTTACACC 5948

Qy    316 eGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspProSe 336
```

```
Db   5949 TTATACAAAAATTAATTCAAGATGGATTAAGTACTTAAAT------GTTAGACCTAAAAC 6002

Qy    336 r---LysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIl 355
Db   6003 CAGAAAAACCCTAGAAGAAAACCTAGGCAATACCATTCAGGAC--------------AT 6047

Qy    355 eGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIl 375
Db   6048 AGGCATGGGCAAG---GACTTCATGTCT------------GAAACACCAAAAGCAATGGC 6092

Qy    375 eArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLys---------IleGlyGl 392
Db   6093 AACAAAAGCCAAAATTGACAAATGGGATCTAATTAAACTAAAGAGCTTCTGCACAGCAAA 6152

Qy    392 nGlyThrValAspPheIleAsnAsnGluIleArgAspProSer---------------- 406
Db   6153 AGAAACTATCATCAGAGTGAAC---------AGGCAACCTACAGAATGGGAGAAAATTTT 6203

Qy    407 ------------------LysAlaLeuIleArgLysValTyrThrGluAlaAspAspLe 420
Db   6204 TGCAACCTACTCATCTGACAAAGGGCTAATATCCAGAATCTACAATGAACTCCAACAAAT 6263

Qy    420 uPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnLysGluIleArgAspPr 440
Db   6264 TTATAAGAAAAAA------------ACAAACAACCCCATCAAAAAGTGGGCCAAGGATAT 6311

Qy    440 oSerLysAlaLeuIleArgLysValSerThrGluAlaAspAsnLeuLeuGluLys 458
Db   6312 GAACAGACACTTCTTAAAAGAAGACACTTATGCAGCCAAAAAACACGTGGAAAAA 6366
```

RESULT 22
ABL04821
ID ABL04821 standard; cDNA; 4443 BP.
XX
AC ABL04821;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 8945.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB60718.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 8945; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4443 BP; 1244 A; 1262 C; 1105 G; 832 T; 0 other;

Alignment Scores:

Pred. No.:	0.6	Length:	4443
Score:	113.50	Matches:	93
Percent Similarity:	33.11%	Conservative:	52
Best Local Similarity:	21.23%	Mismatches:	141
Query Match:	4.81%	Indels:	153
DB:	23	Gaps:	19

US-09-807-459-2 (1-458) x ABL04821 (1-4443)

```
Qy      3 ProSerAspSerValGlyAspValThrLysThrLeuLeuAlaAlaSerGluSerVal--- 21
Db   3322 CCCGATGCCGTTGTAGGGGACTCCGATGCCAAAGTTCTGACATTAGCTGAACACATAATA 3381

Qy     22 ------------------------AspSerAlaAlaAsnAlaTyrMetIleAsnSerAsp 33
Db   3382 GCAGCCATGCCAGAACGAATCAAGAACGAAGCCGATGAAATGATGGTGCTGGGCAGTCCC 3441

Qy     34 MetSerAspTyrLeu------------SerAlaValSerAspAsnPheAlaGluArgIle 49
Db   3442 TTGACAGAACCCCTCACTTCAGAATCTTCAGCGCTGACGGACAGCTTTATGGACCCACTG 3501

Qy     50 CysSerGlnValProLys---------------------------------------- 55
Db   3502 CTGGATTCGCTGCCCAACACACATTTCGACAGCGACTTCAGCTTCGACTTCCATGACCAT 3561

Qy     56 ---------------------GlySerAsnCysSerAlaSer-ValSerAlaTyrMetSe 68
Db   3562 AGCTATCGCTATCACGACGTCAGCACGCCGTGCTCCAGTTTGAGTCCCGCCAGCTCGGGA 3621

Qy     68 rArgCysAla------------------------------------------------- 71
Db   3622 CCGCTGCAGTCCCCGGCCAGTTACTCTATTTTGGGAACCGATCCCTCAGTCAGTTCACCC 3681

Qy     72 ----------------------------LysGlnAspCysLeuThrLeuGlnSerLeuLy 82
Db   3682 AGTCCCCCGCCATCCACCAAACAGTTGACGGAGTTCCTGCACGCCTCCAGCATCTCG-TC 3740

Qy     82 sTyrProLeuGluAlaLysTyrGlnProLeuThrLeuProAspProTyrGlnLeuGluAl 102
Db   3741 GTATCCCTTTGAGGCGGATTTCTCCAAGCTAACCCTAACAGACACGGAACAGCGAGAG-- 3798

Qy    102 aAlaPheIleLeuPheLysGluSerAspAlaAsnProAlaAsnSerThrGluLysArgPh 122
Db   3799 ----------CTTTACGAG--------------GCAGCCAAGTGCATCCAGAAGGCGTA 3833

Qy    122 eTrpMetArgPheArgArgGlyLysAsnHisSerTyrPheHisAspLeuValPheAsnLe 142
Db   3834 T------CGCTCGTACAAGGGTCGACAG-------------------------AAGCT 3860

Qy    142 uLeuGluLysAsnValThrArgAspAlaAspAlaThrAspIleGluAsnPheAlaSerAr 162
Db   3861 CGAAGAACAGAACAAGGAACGGAGCGCT---GCCACTGTCATCCAAAACTACTATAGGCG 3917

Qy    162 gTyrLeuTyrMetAlaThrLeuTyrTyrLysThrTyrThrAsnValAspGluPheGlyAl 182
Db   3918 GTACAAGCAGTACGCC------TACTACAGGCAAATGACAAAT---------------- 3954

Qy    182 aSerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAl 202
Db   3955 ------------------GCTGCCCTGGTGATTCAACATGGCTATCGCTCCTACAGACG 3995

Qy    202 aLeuLysGlnIleIleArgSerAsnLeuProLeuAspIleGlyThrGluHis---SerVa 221
```

```
Db    3996 AAACAAGCGATTCAAGAAGAGCGGCTTGTGCTTGAGCAGCTCCAGCGATCATGGAAGTGT 4055

Qy     221 lSerArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAl 241
           ||||                 ||||||:::  |||:::
Db    4056 GAGCAGCAATTCTCAGTGCCTGTCCAGCTTCTACGATCACTACAAACAGGAT-------- 4107

Qy     241 aLeuProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrVa 261
                                              |||:::   |||      ::
Db    4108 ------------------------------------CAGCAGCAGCTTCACGAGTT 4127

Qy     261 lAlaGlyTyrValAspThrProTrpTyrLysLysTrpTyrMetLysLeuLysAsnPheMe 281
           :              ||||||
Db    4128 GGGCTCCCAACCGAGCACGCCT---------------------------------- 4149

Qy     281 tValAsnArgValPheIleProThrLysLysPhePheAsnLysGluIleArgGluProSe 301
                                              ||||||         |||||
Db    4150 ------------------------------------AAGGAA---ACCAGCCCTTC 4166

Qy     301 rLysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGluAsnLysIleGl 321
           |      ||||||        |||      |||::::::            ||||||
Db    4167 GGGTCCCTTAAAGCGAACCTACTCCCAGTCCACGCAAAATCAAGCTGCCCGGAAAATTCA 4226

Qy     321 yGlnGlyThrValAspPhePheAsnLysGluIleArg---------AspProSerLysAl 338
            |||               |||   ::::::   |||                  :::||
Db    4227 ACAG--------------TTTATGAGACAATCACGCATCAAACTACAGAAAGAGCGAGC 4271

Qy     338 aLeuLysGluLysValSerAsnAspAlaLys---AspLeuPheGluAsnLysIleGlyGl 357
           |   ||||||||:::   :::      :::   :::      :::|||           ||
Db    4272 CGAAAAAGAGAAGCTGGTGCACCAACGCAGGGCGGAATACCTTCAAAACTTGCAGTTTCA 4331

Qy     357 nGlyThrValAspPheIle---------AsnAsnGluIleArgAspProSer 371
           ||||      :::   :::          |||   |||      ||||||
Db    4332 AGGGCAGCAGGAAATGTTGGTCTACCACGAAAATAACATTTCTGCGCCAAGC 4383

RESULT 23
AAV84691
ID   AAV84691 standard; DNA; 10478 BP.
XX
AC   AAV84691;
XX
DT   12-APR-1999  (first entry)
XX
DE   Arabidopsis ESD4 (early short days 4) gene.
XX
KW   ESD4 gene; early short days 4; flowering; transgenic plant; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             3490..6495
FT                   /*tag= a
FT                   /note= "contains introns"
FT   exon            3490..4239
FT                   /*tag= b
FT                   /number= 1
FT   intron          4240..4535
FT                   /*tag= c
FT                   /number= 1
FT   exon            4536..4612
FT                   /*tag= d
FT                   /number= 2
FT   intron          4613..4689
FT                   /*tag= e
FT                   /number= 2
FT   exon            4690..4780
FT                   /*tag= f
FT                   /number= 3
FT   intron          4781..4851
FT                   /*tag= g
FT                   /number= 3
FT   exon            4852..4950
FT                   /*tag= h
FT                   /number= 4
FT   intron          4951..5247
FT                   /*tag= i
FT                   /number= 4
FT   exon            5248..5337
FT                   /*tag= j
FT                   /number= 5
FT   intron          5338..5435
FT                   /*tag= k
FT                   /number= 5
FT   exon            5436..5555
FT                   /*tag= l
FT                   /number= 6
FT   intron          5556..5802
FT                   /*tag= m
FT                   /number= 6
FT   exon            5803..5910
FT                   /*tag= n
FT                   /number= 7
FT   intron          5911..6198
FT                   /*tag= o
FT                   /number= 7
FT   exon            6199..6271
FT                   /*tag= p
FT                   /number= 8
FT   intron          6272..6432
FT                   /*tag= q
FT                   /number= 8
FT   exon            6433..6495
FT                   /*tag= r
FT                   /number= 9
FT   misc_feature    4540..4963
FT                   /*tag= s
FT                   /note= "this region is from ecotype Landsberg
FT                         erecta, the remainder of the sequence is
FT                         from ecotype Columbia"
FT   mutation        2757..3518
FT                   /*tag= t
FT                   /note= "this region is deleted in the eds4 mutant"
XX
PN   WO9856918-A1.
XX
PD   17-DEC-1998.
XX
PF   12-JUN-1998;   98WO-GB01714.
XX
PR   13-JUN-1997;   97GB-0012415.
XX
PA   (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI   Coupland GM;
XX
DR   WPI; 1999-070324/06.
DR   P-PSDB; AAW86184.
XX
PT   New nucleic acid of the Arabidopsis thaliana ESD4 gene - used for
PT   regulating flowering time in plants
XX
PS   Example 1; Fig 2; 73pp; English.
XX
CC   This is the nucleotide sequence of the early short days 4 (ESD4)
CC   gene of Arabidopsis thaliana ecotype Columbia (with part of the
CC   sequence from an ecotype Landsberg erecta clone) that plays a role
CC   in the regulation of flowering time.  Plants carrying a recessive
CC   mutation affecting the ESD4 gene flower earlier than their
CC   wild-types under long and short days.  The esd4 mutant was
CC   identified in a population of gamma-ray mutagenised plants.  The
CC   ESD4 gene was then identified by complementation and sequencing of
CC   the mutant allele.  Regulation of ESD4 gene expression is used to
CC   alter (advance or delay) the time of flowering of transgenic
CC   plants.  Overexpression delays flowering while underexpression
CC   promotes it, e.g. to ensure simultaneous flowering of both parents
CC   in hybrid production, or to regulate flowering according to
CC   climatic conditions.  ESD4 nucleic acids can also be used as
```

```
CC   primers or probes to isolate homologous sequences in other plants
CC   (see also AAV84692-93).
XX
SQ   Sequence 10478 BP; 3195 A; 1770 C; 1986 G; 3527 T; 0 other;

Alignment Scores:
Pred. No.:                1.89          Length:          10478
Score:                    113.50        Matches:         99
Percent Similarity:       38.29%        Conservative:    89
Best Local Similarity:    20.16%        Mismatches:      190
Query Match:              4.81%         Indels:          113
DB:                       20            Gaps:            27

US-09-807-459-2 (1-458) x AAV84691 (1-10478)

Qy      14 LeuLeuAlaAlaSerGluSerValAspSerAlaAlaAsnAlaTyrMetIleAsnSerAsp 33
           :::||||||      |||      |||      ::::::   :::   :::   :::|||
Db    8178 ATCCTTGCAGGAATCGAAGGAGGAGACGATGAATCTGACTATCATGAACTCGTCACGGAT 8237

Qy      34 MetSerAspTyrLeuSerAlaValSerAspAsnPheAlaGluArgIle------------ 49
           :::         |||             |||            |||
Db    8238 CTCAAATCTCTACTTGATACAGACGATGATGAAATTCTCAATCGATTCTACGGTAGTCTC 8297

Qy      50 ----------------------------Cys-----------------SerGlnValPro 54
                                       |||                 |||   |||
Db    8298 TCCTCAATGGCTTCTTCGTTTCTCCGCTGTATCTCCGCCGCCATGGATTCTCCGGTTGAA 8357

Qy      55 LysGlySerAsnCysSerAlaSerValSerAlaTyrMetSerArgCysAlaLysGlnAsp 74
           |||               :::      ||||||:::|||                 :::
Db    8358 TCAGGCCGTCTTGCTATTTTAGCCTCCGACGCTTATCTCAGTCTGCTTCTTTCCACGAAT 8417

Qy      75 CysLeuThrLeuGlnSerLeuLysTyrProLeuGluAlaLysTyrGlnProLeuThrLeu 94
           |||                  |||:::         :::   |||:::
Db    8418 TGC-----------------------CCCGTTTTCACTTTCTTCTCTCCGATTGCGTTT 8453

Qy      95 ProAspProTyrGlnLeuGluAlaAlaPheIleLeuPheLysGluSerAspAlaAsnPro 114
                                     ::::::     ::::::      |||
Db    8454 CTTTCTCTACTTGGCTCAATTCGCCGCTACCTCAAACGCCGTGATGATTCCGCCGGTCAA 8513

Qy     115 AlaAsnSerThrGluLysArgPheTrpMetArgPheArgArgGlyLysAsnHisSerTyr 134
           ::::::::::::::            :::   :::||||||:::
Db    8514 GGGAGTAATTCGCAGCGAGAGAAAGGGAATAAGAAGAAGAGAGGACGTGGT--------- 8564

Qy     135 PheHisAspLeuValPheAsnLeuLeuGluLysAsnVal---ThrArgAspAlaAspAla 153
                                     ::::::|||:::         |||   :::
Db    8565 ----------------------------AAGAGGAATTTAGGGTATGAAGATGGGGAAGAG 8597

Qy     154 ThrAspIleGluAsnPheAlaSerArgTyrLeuTyrMetAlaThrLeuTyrTyrLysThr 173
           |||:::         |||   ::::::   :::::::::
Db    8598 ACTGAAGAAGGTGGATTTGATGCGAAATTGATGTTTATAGTG----------------- 8639

Qy     174 TyrThrAsnValAspGluPheGlyAla------SerPhePheAsnLysLeuSerPheThr 191
                    :::::::::   |||:::      ||||||   :::          |||
Db    8640 ---------TTAGAGAAGCTTGGCTCGGTTCTGAGTTTTGTTCATTTAGATAGATTTCCT 8690

Qy     192 ThrGlyLeuPheGlyTrpGlyIleLysArgAlaLeuLysGlnIleIleArgSerAsnLeu 211
                |||                  :::::::|||   |||   :::   |||   :::
Db    8691 GATAGTTTG-----------------AAATCTTTGGTACAAACTGTG---AGTGAGATT 8729

Qy     212 ProLeuAspIleGlyThrGluHisSer------------ValSerArgLeuGlnHisIle 227
           ||||||   :::      |||||||||                ||||||      :::
Db    8730 CCTTTA---TTGGCGTTGGAGCACTCTGGAGTTTTGAATTATGATCGATTGATGGAAATG 8786

Qy     228 ThrSerSerTyr-----------------LysAspTyrMetAspThrGlnIleProAla 241
                                        |||:::   |||       :::   |||
Db    8787 TGTGGGAAGATTCTGGGAGGAGTGTTGAATTCTGACCATGGAGATATGGCACTCACTGCT 8846

Qy     242 LeuPro-------------------------------------LysPheAlaLysArg 248
                                                       |||:::
Db    8847 GCTGAGATTTCAAAGTCTTTGACACCGTTGCTTCTTATGGGTTTACATCAAGCGAGAAAT 8906

Qy     249 PheSerLeuMetValValGlnArgLeuLeuAlaThrValAlaGlyTyrValAspThrPro 268
           |||:::|||      |||   |||   |||   ::::::|||         |||   |||
```

```
Db    8907 TTTGCGTTGGGATTTGTGTCAAGGAAATTGATGAGTTTGGCT------AAAGATAACCCT 8960

Qy     269 TrpTyrLysLysTrpTyrMetLysLeuLysAsnPheMetValAsnArgValPheIlePro 288
                  ||||||            |||      |||:::|||::::::         |||
Db    8961 GAATTGAAAAAAGTTGTGTCTAATTTGCCTAAGTTTCTGGTTCATAAGGCA------CCT 9014

Qy     289 ThrLysLys------------PhePheAsnLysGluIleArgGluProSerLysAlaLeu 304
              |||               |||      :::   :::   |||      ||||||:::
Db    9015 GAGAAGGCCGAGCCGCGTGGATTTGCAGTGGAGGCGGTACTGGAGATTGTAAAGGCAATG 9074

Qy     305 LysGluLysValSerThrAspThrLysAspLeuPheGluAsnLysIleGlyGlnGlyThr 324
           :::   :::      ::::::      |||   |||      |||:::   ||||||
Db    9075 GAGGTTGAGGGCCAATCAGAGTTTGTTGAT---TTTGTAATGAAGATGTGTCAAGGGAAG 9131

Qy     325 ValAspPhePheAsnLysGluIleArgAspProSerLysAlaLeuLysGluLysValSer 344
             :::|||            :::   |||            |||         :::
Db    9132 TCTAATTTTAGAGTATTAGCTGTT---GATATTATACCTCTGTTGATAAGCTCATTAGGA 9188

Qy     345 AsnAspAlaLysAspLeu--------------PheGluAsnLysIleGlyGlnGlyThr 359
           |||         |||:::                    ::::::      |||   |||
Db    9189 AACCCTCTAGGAGATATTAGTTCAGAGAATGGGTTGAAAGATTCGTGGGGCTTGGGTTGT 9248

Qy     360 ValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValSer 379
           :::|||   :::                |||   |||   ||||||||||||
Db    9249 ATTGATGCTTTAGTTCAGCGGTGTTCAGACACGAGC---GCTTTGATTAGA--------- 9296

Qy     380 ThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsn 399
                |||      |||      :::   :::   |||      |||:::|||::::::
Db    9297 ------GCTCGAGCTTTG------TCCAACTTGGCTCAA---GTTGTGGAGTTCTTGTCT 9341

Qy     400 AsnGluIleArgAspProSerLysAlaLeuIleArgLysValTyrThrGluAlaAspAsp 419
              :::   |||      |||:::::::::::::::::::                  :::
Db    9342 GGTGATGAAAGG------AGTAGGTCGATCCTGAAACAAGCCCTTGGGTTTAACGGTGAG 9395

Qy     420 LeuPheGluAsnLysIleGlyGlnGlyThrVal---AspPheIleAsnLysGluIleArg 438
                |||         |||:::|||   |||   |||   :::   |||
Db    9396 ACTTCAGAG---------GGAAAAGGTGCAGTAACTGACCTTTTGAAGAAAAGATGTGTG 9446

Qy     439 AspProSerLysAlaLeuIleArgLysValSer 449
           |||      ||||||   :::|||:::   :::
Db    9447 GAT---GAGAAGGCGGCTGTAAGGAGAGCAGCT 9476

RESULT 24
AAS76165
ID   AAS76165 standard; cDNA; 4282 BP.
XX
AC   AAS76165;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #11969.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG11978.
XX
```

```
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 11969; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4282 BP; 1735 A; 915 C; 751 G; 881 T; 0 other;

Alignment Scores:
Pred. No.:                0.64          Length:          4282
Score:                    113.00        Matches:         109
Percent Similarity:       34.04%        Conservative:    52
Best Local Similarity:    23.04%        Mismatches:      166
Query Match:              4.79%         Indels:          146
DB:                       23            Gaps:            24

US-09-807-459-2 (1-458) x AAS76165 (1-4282)

Qy      55 LysGlySerAsnCysSerAlaSerValSerAlaTyrMetSerArgCysAlaLysGlnAsp 74
           ||||||   :::               |||     :::      |||   :::   |||
Db    1725 AAAGGGTATTCAATTAGGAAAAGAGGAAGTCAAATTGTCCCTGTTTGCAGACGACATGAT 1784

Qy      75 CysLeuThrLeuGln-----------SerLeuLysTyrProLeuGluAlaLys------ 88
           ||||||:::   :::           |||   |||   |||       |||
Db    1785 TGTTTATCTAGAAAACCCCATCGTCTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTT 1844

Qy      89 TyrGlnProLeuThrLeuProAspPro-TyrGlnLeuGluAlaAlaPheIleLeuPheLy 108
              |||   |||   :::   :::       |||         ||||||:::
Db    1845 CAGCAAAGTCTCAGGATACAAAATCAATGTACAAAAATCACAAGCATTCTTATAC----- 1899

Qy     108 sGluSerAspAlaAsnProAlaAsnSerThrGluLysArgPheTrpMetArgPheArgAr 128
                           |||:::|||:::|||
Db    1900 ------------------ACCAATAACACAGACAAA---------------------- 1917

Qy     128 gGlyLysAsnHisSerTyrPheHisAspLeuValPheAsnLeuLeuGluLysAsnVal-- 147
              :::               :::|||   |||       |||   :::
Db    1918 ----CAAGAGAGCCAAATCATGAGTGAACTCCCATTCACAACTGCTTCAAAGAGAATAAA 1973

Qy     148 ------------------ThrArgAspAlaAspAlaThrAspIleGluAsnPheAlaSe 161
                             |||||||||                 ||||||:::
Db    1974 ATACCTAGGAATCCAACTTACAAGGGATGTGAAGGACCTCTTCAAGGAGAACTACAAACC 2033

Qy     161 r------------------------------------------------------------ 161

Db    2034 ACTGCTCAATGAAATAAAAGAGGATACAAACAAATGGAAGAACGTTCCATGTTCATGGGT 2093

Qy     162 -------------ArgTyrLeuTyrMetAlaThrLeuTyrTyrLysThrTyrThrAsnVa 177
                          :::   ||||||   |||         |||
Db    2094 AGGAGGAAGAATCAATCTCGTGAAAATGGCCATACTGCCCAAGGTAATTTAC-------- 2145

Qy     177 lAspGluPheGlyAla--------------------SerPhePheAsnLysLeuSerPh 190
                  |||   |||                   :::||||||   :::|||
Db    2146 ----AGATTCAATGCCATCCCCATCAAGCTACCAATGACTTTCTTCACAGAATTGGAAAA 2201

Qy     190 eThrThrGlyLeuPheGlyTrpGlyIleLysArgAlaLeuLysGlnIleIleArgSerAs 210
           ||||||      |||   |||      ||||||||||||      |||   :::|||::
Db    2202 AACTACTTTAAAGTTCATATGGAACCAAAAGAGAGCCCTC------ATTGCCAAGTCATC 2255

Qy     210 nLeu--ProLeuAspIleGlyThrGluHisSerValSerArgLeuGlnHisIleThrSer 229
           :|||  |||   :::             |||       :::   ||||||   |||
Db    2256 CCTAAGCCAAAAGAACAAAACTGGAGGCATCACGCTACCTGACTTCAAACTATACTACAA 2315

Qy     230 SerTyrLysAspTyrMetAspThrGlnIleProAlaLeuProLysPheAlaLysArgPhe 249
              |||   :::             :::      :::||||||      :::|||:::
Db    2316 GGCTACAGTAAC------CAAAACATCATGGTACTGGTACCAAAA------CAGAGATAT 2363

Qy     250 SerLeuMetValValGlnArgLeuLeu---------AlaThr-ValAlaGlyTyrVal-- 265
              |||         |||   |||         |||||| :::      |||:::
Db    2364 AGATCAATGGAACAGAACAGAGCCCTCAGAAATAATGCCACATATCTACAACTATCTGAT 2423

Qy     266 ----AspThrPro---------------Trp--------------TyrLysLysTrp-- 273
               |||   |||               |||              :::   ||||||
Db    2424 CTTTGACAAACCTGACAAAAACAAGCAATGGGGAAAGGATTCCCTATTTAATAAATGGTG 2483

Qy     274 ---------------------------TyrMetLysLeuLysAsnPheMetValAsnAr 284
                                        :::|||||      |||:::
Db    2484 CTGGGAAAACTGGCTAGCCATATGTAGAAAGCTAAAACTGGATCCCTTCCTTACGCCTTA 2543

Qy     284 gValPheIleProThrLysLysPhePheAsnLysGluIleArgGluProSer---LysAl 303
              |||   ::::::         :::      :::|||      :::   |||
Db    2544 TACAAAAATTGATTCAAGATGGATTAAAGACTTACATGTTAGACCTAAAACCATAAAAAC 2603

Qy     303 aLeuLysGluLysValSerThrAspThrLysAspLeuPheGluAsnLysIleGlyGlnGl 323
           |||:::|||   :::             :::|||                ||||||   ||
Db    2604 CCTAGAAGAAAACCTAGGCAATACCATTCAGGAC--------------ATAGGCATGGG 2648

Qy     323 yThrValAspPhePheAsnLysGluIleArgAspProSerLysAlaLeuLysGluLysVa 343
           |      ||||||   :::|||                ||||||:::      |||
Db    2649 CAAG---GACTTCATGTCTAAA------------ACACAAAAAGCAATGGCAACAAAAGC 2693

Qy     343 lSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIl 363
              |||      ||||||   :::   |||
Db    2694 CAAAATTGACAAATGGGATCTAATTAAACTAAAG------------------------- 2727

Qy     363 eAsnAsnGluIleArgAspProSerLysAlaLeuIleArg------LysValSerThrGl 381
                        |||         :::      ||||||      :::      |||
Db    2728 -------AGCTTCCGCACAGCAAAAGAAACTACCATCAGAGTGAACAGGCAACCTACAAC 2780

Qy     381 yAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGl 401
              |||   :::|||
Db    2781 ATGGGAGAAAATTTTT----------------------------------------GC 2798

Qy     401 uIleArgAspProSerLysAlaLeuIleArgLysValTyrThrGluAlaAspAspLeuPh 421
           |||            |||   ||||||   :::|||||||   |||         :::::
Db    2799 AATCTACTCATCTGACAAAGGGCTAATATCCAGAGTCTACAATGAACTCAAACAAATTTA 2858

Qy     421 eGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnLysGluIleArgAspProSe 441
           ::::   |||                 |||      |||:::    ::
Db    2859 CAAGAAAAAA----------------------ACAAACAACCCCATCAAAAAGTGGGC 2894

Qy     441 rLysAlaLeuIleArgLysValSerThrGlu 451
           :|||   :::   |||      |||   |||
Db    2895 AAAGGATATGAACAGACACTTCTCCAAAGAA 2925

RESULT 25
AAS74637
ID   AAS74637 standard; cDNA; 11087 BP.
XX
AC   AAS74637;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #10441.
```

```
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG10450.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 10441; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 11087 BP; 4422 A; 2400 C; 2076 G; 2189 T; 0 other;

Alignment Scores:
Pred. No.:              2.28          Length:         11087
Score:                  113.00        Matches:        109
Percent Similarity:     34.67%        Conservative:   55
Best Local Similarity:  23.04%        Mismatches:     179
Query Match:            4.79%         Indels:         130
DB:                     23            Gaps:           22

US-09-807-459-2 (1-458) x AAS74637 (1-11087)

Qy      55 LysGlySerAsnCysSerAlaSerValSerAlaTyrMetSerArgCysAlaLysGlnAsp 74
           ||||||   :::   |||         |||      :::      |||   :::   |||
Db    9544 AAAGGGTATTCAATTAGCAAAAGAGGAAGTCAAATTGTCCCTGTTTGCAGACGACATGAT 9603

Qy      75 CysLeuThrLeuGln-----------SerLeuLysTyrProLeuGluAlaLysTyr--- 89
           ||||||:::   :::           |||   |||  |||         |||
Db    9604 TGTTTATCTAGAAAACCCCATCGTCTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTT 9663

Qy      90 ---GlnProLeuThrLeuProAspPro-TyrGlnLeuGluAlaAlaPheIleLeuPheLy 108
              |||   |||   :::   :::       |||         |||||||:::

Db    9664 CGGCAAAGTCTCAGGATACAAAATCAATGTACAAAAATCACAAGCATTCTTATACACCAA 9723

Qy     108 sGluSerAspAlaAsnProAlaAsnSerThrGluLysArgPheTrpMetArgPheArgAr 128
                                      :::|||      |||   :::      :::||
Db    9724 CACCAGACAAACAGAGAGCCAAATCATGAGTGAACTCCCATTCACAATTGCTTCAAAGAG 9783

Qy     128 gGlyLysAsnHisSerTyrPheHisAspLeuValPheAsnLeuLeuGluLysAsnValTh 148
           |   |||                                        |||         :::||
Db    9784 AATAAAA--------------------------------TACCTAGGAATCCAACTTAC 9810

Qy     148 rArgAspAlaAspAlaThrAspIleGluAsnPheAlaSer---------------------- 161
           |||||||                |||||||:::
Db    9811 AAGGGATGTGAAGGACCTCTTCAAGGAGAACTACAAACCACTGCTCAAGGAAATAAAAGA 9870

Qy     162 ----------------------------------------------ArgTyrLeuTy 165
                                                           :::
Db    9871 GGACACAAACAAATGGAAGAACATTCCATGCTCATGGGTAGGAAGAATCAATATCGTGAA 9930

Qy     165 rMetAlaThrLeuTyrTyrLysThrTyrThrAsnValAspGluPheGlyAla-------- 182
           ||||||   |||           |||               |||   |||
Db    9931 AATGGCCATACTGCCCAAGGTAATTTAC-----------AGATTCAATGCCATCCCCAT 9978

Qy     183 -------------SerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGl 198
                        :::|||||||   :::|||      ||||||      |||   |||
Db    9979 CAAGCTACCAATGACTTTCTTCACAGAATTGGAAAAAACTACTTTAAAGTTCAAATGGAA 10038

Qy     198 yIleLysArgAlaLeuLysGlnIleIleArgSerAsnLeu--ProLeuAspIleGlyThr 217
           ||||||||||         |||   :::|||   |||  |||   :::
Db   10039 CCAAAAAAGAGCCTGC------ATTGCCAAGTCAATCCTAAGCCAAAAGAACAAAGCTGG 10092

Qy     218 GluHisSerValSerArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThr 237
           |||      :::||||||||||   |||         |||   :::
Db   10093 AGGCATCATGCTACCCGACTTCAAACTATACTACAAGGCTACAGTAAT------CAAAAC 10146

Qy     238 GlnIleProAlaLeuProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeu 257
           :::      :::||||||      :::|||:::      |||           |||
Db   10147 AGCATGGTACTGGTACCAAAA------CAGAGATATAGACCAATGGAACAGAACAGAGCC 10200

Qy     258 LeuAlaThrValAlaGlyTyr---------------Val-AspThrPro----------- 268
           |||         |||   |||               ::: |||   |||
Db   10201 CTCAGAAATAATGCCGCATATCTACAACTATCTGATCTTCGACAAACCTGACCAAAACAA 10260

Qy     269 ----Trp---------------TyrLysLysTrpTyr---------------------- 274
               |||               :::   |||||||||
Db   10261 GCAATGGGGAAAGGATTCCCTATTTAATAAATGGTACTGGGAAAACTGCCTAGCCATATG 10320

Qy     275 -------MetLysLeuLysAsnPheMetValAsnArgValPheIleProThrLysLysPh 292
                ::::||||||      |||:::               |||   ::::::
Db   10321 TAGAAAGCTGAAACTAGATCCCTTCCTTACACCTTATACAAAAATTAATTCAAGATGGAT 10380

Qy     292 ePheAsnLysGluIleArg---GluProSerLysAlaLeuLysGluLysValSerThrAs 311
              :::      :::|||   :::      ||||||||||:::|||   :::
Db   10381 TAAAGACTTAAATGTTAGACTTAAAACCATAAAAGCCCTAGAAGAAAACCTAGGCAATAC 10440

Qy     311 pThrLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGl 331
              :::|||               ||||||   |||      ||||||   :::|||
Db   10441 CATTCAGGAC--------------ATAGGCATGGGCAAG---GACTTCATGTCTAAA-- 10480

Qy     331 uIleArgAspProSerLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPh 351
                        ||||||:::      |||         |||      ||||||
Db   10481 ----------ACACCAAAAGCAATGGCAACAAAAGACAAAATTGACAAATGGGATCTAAT 10530

Qy     351 eGluAsnLys---------IleGlyGlnGlyThrValAspPheIleAsnAsnGluIleAr 368
              :::   |||               :::   |||         :::|||          ||
Db   10531 TAAACCAAAGAGCTTCTGCACAGCAAAAGAAACTACCATCAGAGTGAAC---------AG 10581

Qy     368 gAspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAs 388
           |   |||:::|||                        |||   :::|||
Db   10582 GCAACCTACAAAA--------------------------TGGGAGAAAATTTTCGCAAC 10614

Qy     388 nLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAl 408
                                                              |||
Db   10615 C--------------------------------------TACTCATCTGACAAAGG 10632
```

```
Qy    408 aLeuIleArgLysValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGl 428
          ||||||   :::   |||   |||         :::::::::   |||
Db  10633 GCTAATATCCAGAAACTACAATGAACTCAAACAAATTTACAAGAAAAAA---------- 10681

Qy    428 yThrValAspPheIleAsnLysGluIleArgAspProSerLysAlaLeuIleArgLysVa 448
          |||:::      |||   |||         |||   ::::::         ::::::
Db  10682 -ACAATCGGCCCCATCAAAAAGTGGGCAGAGGATATGAACAGGCACTTCTCAAAAGAAGA 10740

Qy    448 lSerThrGluAlaAspAsnLeuLeuGluLys 458
          :::      |||        ::::::|||
Db  10741 CACTTATGCAGCCAAAAAACACATGAAAAAA 10771

RESULT 26
AAK68993
ID   AAK68993 standard; DNA; 2375 BP.
XX
AC   AAK68993;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23805.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
```

```
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 23805; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 2375 BP; 941 A; 489 C; 444 G; 501 T; 0 other;

Alignment Scores:
Pred. No.:              0.326          Length:          2375
Score:                  112.50         Matches:         74
Percent Similarity:     34.44%         Conservative:    51
Best Local Similarity:  20.39%         Mismatches:      131
Query Match:            4.77%          Indels:          107
DB:                     22             Gaps:            16

US-09-807-459-2 (1-458) x AAK68993 (1-2375)

Qy     174 TyrThrAsnValAspGluPheGlyAlaSerPhePheAsnLysLeuSerPheThrThrGly 193
           |||   |||      :::   |||:::        ::::::|||||||||||||||||
Db     538 TACATCAATAACAGACAAACAGGGAGCCAAATCATGAGTGAACTCTCATTCACAACTGCT 597

Qy     194 --------------LeuPheGlyTrpGlyIleLysArgAlaLeuLysGlnIleIleArg 208
                         |||      :::   |||   :::|||   :::   :::
Db     598 TCAAAGAGAATAAAATACCTAGGAATCCAACTTACAAGGGATGTGAAGGGCCTCTTCAAG 657

Qy     209 SerAsnLeu---ProLeuAspIleGlyThrGluHisSerValSerArgLeuGlnHisIle 227
              |||      ||||||           :::         ::::::   ::::::|||
Db     658 GAGAACTACAAACCACTGCACAAGGAAATAAAAGAGGATACAAACAAATGGAAGAACATT 717

Qy     228 ThrSerSerTyrLysAspTyrMetAsp---ThrGlnIleProAlaLeuProLysPheAla 246
                 |||:::         ::::::     ::::::      |||||||||
Db     718 CCATGCTCATGGGTAGGAAGAATCAATATCGTGAAAATGGCCATACTGCCCAAGGTAATT 777

Qy     247 LysArgPheSerLeuMetValValGln--------------------ArgLeuLeuAla 259
           ||||||:::         ::::::                          |||
Db     778 TATAGATTCAATGCGTTCCCCATTAAGCTACCAATGACATTCTTCACAGAATTGGAAAAA 837

Qy     260 ThrValAlaGlyTyrVal---------------------------------------- 265
           |||         ::::::
Db     838 ACTGCTTTAAAGTTCATATGAAACAAAAAAAGAGCCTGCATTGCCAAGTCAAACCTAAGC 897

Qy     265 ---------------------------------------------------------- 265

Db     898 CAAAAGAACAAAGCTGGAGGCATCACGCTACCTGACTTCAAACTATACTACAAGGCTACA 957

Qy     266 ------AspThrProTrpTyrLysLysTrpTyrMetLys-------------------- 276
                   |||   ||||||     ||||||
Db     958 CTAACCAAAACAGCATGGTAC------TGGTACCAAAACAGAGATATAGACCAATGGAAC 1011

Qy     277 ----------------------------LeuLysAsnPheMetValAsnArgValPhe 286
                                        :::   |||:::::::::
Db    1012 AGAACAGAGCCCTCAGAAATAACACCACACATCTACAACTATCTGATCTTTGACAAACCT 1071

Qy     287 IleProThrLysLysPhe--------------PheAsnLysGluIleArgGluProSer 301
                 |||||||:::               |||||||||         |||
Db    1072 GACAAAAACAAGAAATGGGGAAAGGATTCCCTATTTAATAAACGGTGCTGGGAAAACTGG 1131

Qy     302 LysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGluAsnLysIleGly 321
           |||:::      |||:::      |||                        ||||||
Db    1132 CTAGCCATATGTAGAAAGCTGAAACTGGATCCCTTCCTTACACCTTATACAAAAATTAAT 1191

Qy     322 GlnGlyThrValAspPhePheAsnLysGluIleArgAspProSer---LysAlaLeuLys 340
                 :::         |||      :::|||      :::   |||   |||:::
Db    1192 TCAAGATGGATTAAAGACTTAAAC------GTTAGACCTAAAACCATAAAAACCCTAGAA 1245

Qy     341 GluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThrVal 360
           |||   :::   |||      :::|||                ||||||   |||
Db    1246 GAAAACCTAGGCAATATCATTCAGGAC--------------ATAGGCATGGGAAAG--- 1287

Qy     361 AspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValSerThr 380
           ||||||::::::                        ||||||:::      |||
Db    1288 GACTTCATGTCTAAA-----------ACACCAAAAGCAATGGCAACAAAAGCCAAAATT 1335

Qy     381 GlyAlaGluAspLeuPheGluAsnLys---------IleGlyGlnGlyThrValAspPhe 397
                 ||||||   :::   |||                :::   |||
Db    1336 GACAAATGGGATCTAATTAAACTAAAGAGCTTCTGCACAGCAAAAGAAACTACCATCAGA 1395

Qy     398 IleAsnAsnGlu--------------------------IleArgAspProSerLysAla 408
           :::|||   :::                          |||      |||   |||
Db    1396 GTGAACAGGCAACCTACAGAATGGGAGAAAGTTTTTGCAATCTACTCACCTGACAAAGGG 1455

Qy     409 LeuIleArgLysValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGly 428
           ||||||   :::::|||   |||         :::::::::   |||
Db    1456 CTAATATCCAGAATCTACAATGAACTCAAACAAATTTACAAGAAAAAA------------ 1503

Qy     429 ThrValAspPheIleAsnLysGluIleArgAspProSerLysAlaLeuIleArgLysVal 448
                       |||      |||:::      :::|||   :::   |||
Db    1504 ------------ACAAACAACCCCATCAAAAAGTGGGCAAAGGATATGAACAGACACTTC 1551

Qy     449 SerThrGlu 451
           |||   |||
Db    1552 TCAAAAGAA 1560

RESULT 27
AAK68992
ID   AAK68992 standard; DNA; 42738 BP.
```

XX
AC AAK68992;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23804.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.

```
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 23804; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 42738 BP; 15539 A; 8490 C; 8408 G; 10301 T; 0 other;

Alignment Scores:
Pred. No.:                15.5          Length:         42738
Score:                    112.50        Matches:        74
Percent Similarity:       34.44%        Conservative:   51
Best Local Similarity:    20.39%        Mismatches:     131
Query Match:              4.77%         Indels:         107
DB:                       22            Gaps:           16

US-09-807-459-2 (1-458) x AAK68992 (1-42738)

Qy     174 TyrThrAsnValAspGluPheGlyAlaSerPhePheAsnLysLeuSerPheThrThrGly 193
           |||   |||      :::   |||:::         ::::::|||||||||||||||||
Db   40901 TACATCAATAACAGACAAACAGGGAGCCAAATCATGAGTGAACTCTCATTCACAACTGCT 40960

Qy     194 --------------LeuPheGlyTrpGlyIleLysArgAlaLeuLysGlnIleIleArg 208
                        |||      :::   |||   :::|||   :::   :::
Db   40961 TCAAAGAGAATAAAATACCTAGGAATCCAACTTACAAGGGATGTGAAGGGCCTCTTCAAG 41020

Qy     209 SerAsnLeu---ProLeuAspIleGlyThrGluHisSerValSerArgLeuGlnHisIle 227
           |||      ||||||           :::         ::::::   ::::::|||
Db   41021 GAGAACTACAAACCACTGCACAAGGAAATAAAAGAGGATACAAACAAATGGAAGAACATT 41080

Qy     228 ThrSerSerTyrLysAspTyrMetAsp---ThrGlnIleProAlaLeuProLysPheAla 246
             |||:::         ::::::      ::::::      |||||||||
Db   41081 CCATGCTCATGGGTAGGAAGAATCAATATCGTGAAAATGGCCATACTGCCCAAGGTAATT 41140

Qy     247 LysArgPheSerLeuMetValValGln--------------------ArgLeuLeuAla 259
           ||||||:::         ::::::                        |||
Db   41141 TATAGATTCAATGCGTTCCCCATTAAGCTACCAATGACATTCTTCACAGAATTGGAAAAA 41200

Qy     260 ThrValAlaGlyTyrVal---------------------------------------- 265
           |||         ::::::
```

```
Db   41201 ACTGCTTTAAAGTTCATATGAAACAAAAAAAGAGCCTGCATTGCCAAGTCAAACCTAAGC 41260

Qy     265 ---------------------------------------------------------- 265

Db   41261 CAAAAGAACAAAGCTGGAGGCATCACGCTACCTGACTTCAAACTATACTACAAGGCTACA 41320

Qy     266 ------AspThrProTrpTyrLysLysTrpTyrMetLys---------------------- 276
                 |||   ||||||      ||||||
Db   41321 CTAACCAAAACAGCATGGTAC------TGGTACCAAAACAGAGATATAGACCAATGGAAC 41374

Qy     277 ------------------------------LeuLysAsnPheMetValAsnArgValPhe 286
                                           :::   |||:::::::::
Db   41375 AGAACAGAGCCCTCAGAAATAACACCACACATCTACAACTATCTGATCTTTGACAAACCT 41434

Qy     287 IleProThrLysLysPhe---------------PheAsnLysGluIleArgGluProSer 301
                 ||||||:::                 |||||||||         |||
Db   41435 GACAAAAACAAGAAATGGGGAAAGGATTCCCTATTTAATAAACGGTGCTGGGAAAACTGG 41494

Qy     302 LysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGluAsnLysIleGly 321
           |||:::      |||:::      |||                         ||||||
Db   41495 CTAGCCATATGTAGAAAGCTGAAACTGGATCCCTTCCTTACACCTTATACAAAAATTAAT 41554

Qy     322 GlnGlyThrValAspPhePheAsnLysGluIleArgAspProSer---LysAlaLeuLys 340
                 :::       |||      :::|||      :::   |||   |||:::
Db   41555 TCAAGATGGATTAAAGACTTAAAC------GTTAGACCTAAAACCATAAAAACCCTAGAA 41608

Qy     341 GluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThrVal 360
           |||   :::   |||      :::|||               ||||||   |||
Db   41609 GAAAACCTAGGCAATATCATTCAGGAC--------------ATAGGCATGGGAAAG--- 41650

Qy     361 AspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValSerThr 380
           ||||||:::::                     ||||||:::      |||
Db   41651 GACTTCATGTCTAAA------------ACACCAAAAGCAATGGCAACAAAAGCCAAAATT 41698

Qy     381 GlyAlaGluAspLeuPheGluAsnLys---------IleGlyGlnGlyThrValAspPhe 397
                 ||||||   :::   |||           :::   |||
Db   41699 GACAAATGGGATCTAATTAAACTAAAGAGCTTCTGCACAGCAAAAGAAACTACCATCAGA 41758

Qy     398 IleAsnAsnGlu--------------------------IleArgAspProSerLysAla 408
           :::|||   :::                         |||      |||   |||
Db   41759 GTGAACAGGCAACCTACAGAATGGGAGAAAGTTTTTGCAATCTACTCACCTGACAAAGGG 41818

Qy     409 LeuIleArgLysValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGly 428
           ||||||   ::::::|||   |||         :::::::   |||
Db   41819 CTAATATCCAGAATCTACAATGAACTCAAACAAATTTACAAGAAAAAA------------ 41866

Qy     429 ThrValAspPheIleAsnLysGluIleArgAspProSerLysAlaLeuIleArgLysVal 448
                       |||      |||:::      :::|||   :::   |||
Db   41867 ------------ACAAACAACCCCATCAAAAAGTGGGCAAAGGATATGAACAGACACTTC 41914

Qy     449 SerThrGlu 451
           |||   |||
Db   41915 TCAAAAGAA 41923
```

```
RESULT 28
AAX35720
ID   AAX35720 standard; cDNA; 1422 BP.
XX
AC   AAX35720;
XX
DT   09-JUL-1999  (first entry)
XX
DE   cDNA encoding a protein identified by the signal sequence trap method.
XX
KW   Signal sequence trap method; SST method; immunisation; inhibition;
KW   infection; allergy; cancer; regulation; tissue formation; tissue repair;
KW   activin activity; inhibin activity; chemokine activity;
KW   cytokine activity; blood coagulation regulation; agonist; antagonist;
KW   metabolic disorder; hormonal disorder; immune disorder;
KW   severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer;
KW   wound; ss.
XX
OS   Homo sapiens.
```

```
XX
PN   WO9918126-A1.
XX
PD   15-APR-1999.
XX
PF   06-OCT-1998;   98WO-JP04514.
XX
PR   07-OCT-1997;   97JP-0274674.
XX
PA   (ONOY ) ONO PHARM CO LTD.
XX
PI   Fukushima D,  Shibayama S,  Tada H;
XX
DR   WPI; 1999-277254/23.
DR   P-PSDB; AAY02371.
XX
PT   Polypeptides identified by the signal sequence trap method from a
PT   human cDNA library
XX
PS   Claim 5; Page 139-140; 281pp; Japanese.
XX
CC   AAX35694-X35747 represent cDNA sequences that encode novel polypeptides
CC   (AAY02358-84) which are identified from a human placental cDNA library
CC   by the signal sequence trap (SST) method. The polypeptides have a
CC   broad range of physiological activity, including immunisation against
CC   and inhibition of infections, allergies and cancer; regulation of tissue
CC   formation and repair; activin/inhibin activity; chemokine/cytokine
CC   activity; blood coagulation regulation; and receptor/ligand agonist
CC   or antagonist activity. The polypeptides can be used for prevention
CC   and treatment of disorders including infections by bacteria, yeasts and
CC   viruses (including HIV) and protozoa; metabolic and hormonal disorders;
CC   immune disorders (including severe combined immunodeficiency (SCID)
CC   and AIDS; thrombosis; cancer; and traumatic or surgical wounds.
XX
SQ   Sequence 1422 BP; 538 A; 213 C; 332 G; 339 T; 0 other;

Alignment Scores:
Pred. No.:              0.231          Length:          1422
Score:                  111.00         Matches:         95
Percent Similarity:     37.42%         Conservative:    79
Best Local Similarity:  20.43%         Mismatches:      165
Query Match:            4.71%          Indels:          126
DB:                     20             Gaps:            26

US-09-807-459-2 (1-458) x AAX35720 (1-1422)

Qy      62 SerValSerAlaTyrMetSerArgCysAlaLysGlnAspCysLeuThrLeuGlnSerLeu 81
           :::::      |||   |||      :::      |||      :::          :::
Db     181 ACCATAAAATGCTATGAGAGTTTTATGAGTGGTGCCGAC------AGCTTTGATGAGATG 234

Qy      82 LysTyrProLeuGluAlaLysTyrGlnProLeuThrLeuProAspProTyrGlnLeuGlu 101
              |||::::::|||   :::   |||         |||   ::::::||||||
Db     235 AATGCAGAGCTGCAGTCAAAACTGAAGGATTTATTTAATGTGGATGCTTTTAAGCTGGAA 294

Qy     102 -------------------------------------AlaAlaPheIleLeuPheLys 108
                                                  |||          :::
Db     295 TCATTAGAAGCAAAAAACAGAGCATTGAATGAACAGATTGCAAGATTGGAACAAGAAAGA 354

Qy     109 GluSerAspAlaAsnProAlaAsnSerThrGluLys------------------------ 120
           |||   :::   |||         |||      |||
Db     355 GAAAAAGAACCGAATCGTCTAGAGTCGTTGAGAAAACTGAAGGCTTCCTTACAAGGAGAT 414

Qy     121 ------ArgPheTrpMetArgPheArgArgGlyLysAsnHisSerTyrPheHisAspLeu 138
              ::::::                  ::::::||||||         |||
Db     415 GTTCAAAAGTATCAGGCATACATGAGCAATTTGGAGTCTCATTCAGCCATTCTTGACCAG 474

Qy     139 ValPheAsnLeuLeuGluLysAsnValThrArg----------------------Asp 150
           |||   |||   :::   :::   |||
Db     475 AAATTAAATGGTCTCAATGAGGAAATTGCTAGAGTAGAACTAGAATGTGAAACAATAAAA 534

Qy     151 AlaAspAlaThrAspIleGluAsnPheAlaSerArgTyrLeuTyrMetAlaThrLeuTyr 170
              :::   |||   ::::::|||                 |||
Db     535 CAGGAGAACACTCGACTACAGAATATCATTGACAACCAGAAGTAC--------------- 579

Qy     171 TyrLysThrTyrThrAsnValAspGluPheGlyAlaSerPhePheAsnLysLeuSerPhe 190
               :::      :::::::::
Db     580 ------TCAGTTGCAGACATTGAGCGAATAAATCATGAA---------------------- 612

Qy     191 ThrThrGlyLeuPheGlyTrpGlyIleLysArgAlaLeuLysGlnIleIleArgSerAsn 210
                                        :::      |||:::|||   |||   :::
Db     613 --------------------------AGAAATGAATTGCAGCAGACTATT---AATAAA 642

Qy     211 LeuProLeuAspIleGlyThrGluHisSerValSerArgLeuGlnHisIleThrSerSer 230
           |||      |||:::      |||           :::|||   :::
Db     643 TTAACCAAGGACCTGGAAGCTGAACAA------CAGAAGTTGTGGAATGAGGAGTTAAAA 696

Qy     231 Tyr---------LysAspTyrMetAspThrGlnIleProAlaLeuProLysPheAlaLys 247
           |||         |||:::   ::::::|||||:::            |||   |||:::
Db     697 TATGCCAGAGGCAAAGAAGCGATTGAAACACAATTAGCAGAGTATCACAAATTGGCTAGA 756

Qy     248 ArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValAlaGlyTyrValAspThr 267
           :::      |||   ::::::   :::                  ||||||
Db     757 AAA------TTAAAACTTATTCCTAAAGGTGCTGAGAATTCCAAAGGTTAT--------- 801

Qy     268 ProTrpTyrLysLysTrpTyrMetLysLeuLys--------------AsnPheMetVal 282
                         :::   :::|||                    |||   :::|||
Db     802 ------------GACTTTGAAATTAAGTTTAATCCCGAGGCTGGTGCCAACTGCCTTGTC 849

Qy     283 Asn---------ArgValPheIleProThrLysLysPheAsnLysGluIleArgGlu 299
                    :::|||:::::::|||   |||:::      |||:::         |||
Db     850 AAATACAGGGCTCAAGTTTATGTACCTCTTAAGGAACTCCTGAATGAAACTGAAGAAGAA 909

Qy     300 ProSerLysAlaLeuLysGluLysValSer---ThrAspThrLysAsp------------ 314
              :::|||||||||   :::|||:::          ||||||   :::
Db     910 ATTAATAAAGCCCTAAATAAAAAAATGGGTTTGGAGGATACTTTAGAACAATTGAATGCA 969

Qy     315 ---LeuPheGluAsnLysIleGlyGlnGlyThrVal----------------------- 325
              :::   |||:::|||        ||||||:::
Db     970 ATGATAACAGAAAGCAAGAGAAGTGTGGGAACTCTGAAAGAAGAAGTTCAAAAGCTGGAT 1029

Qy     326 AspPhePheAsnLysGluIleArgAspProSerLysAlaLeuLysGluLysValSerAsn 345
             |||   :::   ::::::|||::::::      :::          ||||||   ::::::
Db    1030 GATCTTTACCAACAAAAAATTAAGGAAGCAGAGGAAGAG---GATGAAAAATGTGCCAGT 1086

Qy     346 AspAla-----------------LysAspLeuPheGluAsnLysIleGlyGlnGlyThr 359
                               :::      |||   |||   |||:::   :::   ||||||
Db    1087 GAGCTTGAGTCCTTGGAGAAACACAAGCACCTGCTAGAAAGTACTGTTAACCAGGGGCTC 1146

Qy     360 ValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValSer 379
           :::   :::   |||||||:::          :::                  |||
Db    1147 AGTGAAGCTATG---AATGAATTAGATGCTGTTCAGCGGGAATACCAACTAGTTGTGCAA 1203

Qy     380 ThrGlyAlaGluAspLeuPheGluAsnLysIleGly---------GlnGlyThrValAsp 396
           |||      |||:::            |||:::|||         |||      ::::::
Db    1204 ACCACGACTGAAGAA------AGACGAAAAGTGGGAAATAACTTGCAACGTCTGTTAGAG 1257

Qy     397 PheIleAsnAsnGluIleArgAspProSerLysAlaLeu---------IleArgLysVal 413
              :::         :::              |||   |||         |||   ||||||
Db    1258 ATGGTTGCTACACATGTTGGGTCTGTAGAGAAACATCTTGAGGAGCAGATTGCTAAAGTT 1317

Qy     414 ---------TyrThrGlu------AlaAspAspLeuPheGluAsnLysIleGlyGlnGly 428
                   |||   |||      ::::::||||||   ||||||
Db    1318 GATAGAGAATATGAAGAATGCATGTCAGAAGATCTCTCGGAAAAT--------------- 1362

Qy     429 ThrValAspPheIleAsnLysGluIleArgAsp------ProSerLysAlaLeuIleArg 446
                          ||||||||||||||||            ||||||   :::
Db    1363 --------------ATTAAAGAGATTAGAGATAAGTATGAGAAGAAAGCTACTCTAATT 1407

Qy     447 LysValSerThrGlu 451
           |||   |||   |||
Db    1408 AAGTCTTCTGAAGAA 1422

RESULT 29
AAX35721
ID   AAX35721 standard; cDNA; 1613 BP.
XX
```

```
AC   AAX35721;
XX
DT   09-JUL-1999  (first entry)
XX
DE   cDNA encoding a protein identified by the signal sequence trap method.
XX
KW   Signal sequence trap method; SST method; immunisation; inhibition;
KW   infection; allergy; cancer; regulation; tissue formation; tissue repair;
KW   activin activity; inhibin activity; chemokine activity;
KW   cytokine activity; blood coagulation regulation; agonist; antagonist;
KW   metabolic disorder; hormonal disorder; immune disorder;
KW   severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer;
KW   wound; ss.
XX
OS   Homo sapiens.
XX
PN   WO9918126-A1.
XX
PD   15-APR-1999.
XX
PF   06-OCT-1998;   98WO-JP04514.
XX
PR   07-OCT-1997;   97JP-0274674.
XX
PA   (ONOY ) ONO PHARM CO LTD.
XX
PI   Fukushima D,  Shibayama S,  Tada H;
XX
DR   WPI; 1999-277254/23.
DR   P-PSDB; AAY02371.
XX
PT   Polypeptides identified by the signal sequence trap method from a
PT   human cDNA library
XX
PS   Claim 4; Page 141-144; 281pp; Japanese.
XX
CC   AAX35694-X35747 represent cDNA sequences that encode novel polypeptides
CC   (AAY02358-84) which are identified from a human placental cDNA library
CC   by the signal sequence trap (SST) method. The polypeptides have a
CC   broad range of physiological activity, including immunisation against
CC   and inhibition of infections, allergies and cancer; regulation of tissue
CC   formation and repair; activin/inhibin activity; chemokine/cytokine
CC   activity; blood coagulation regulation; and receptor/ligand agonist
CC   or antagonist activity. The polypeptides can be used for prevention
CC   and treatment of disorders including infections by bacteria, yeasts and
CC   viruses (including HIV) and protozoa; metabolic and hormonal disorders;
CC   immune disorders (including severe combined immunodeficiency (SCID)
CC   and AIDS; thrombosis; cancer; and traumatic or surgical wounds.
XX
SQ   Sequence 1613 BP; 624 A; 243 C; 359 G; 387 T; 0 other;

Alignment Scores:
Pred. No.:              0.274          Length:          1613
Score:                  111.00         Matches:         95
Percent Similarity:     37.42%         Conservative:    79
Best Local Similarity:  20.43%         Mismatches:      165
Query Match:            4.71%          Indels:          126
DB:                     20             Gaps:            26

US-09-807-459-2 (1-458) x AAX35721 (1-1613)

Qy      62 SerValSerAlaTyrMetSerArgCysAlaLysGlnAspCysLeuThrLeuGlnSerLeu 81
           ::::::      |||   |||      :::      |||      :::         :::
Db     279 ACCATAAAATGCTATGAGAGTTTTATGAGTGGTGCCGAC------AGCTTTGATGAGATG 332

Qy      82 LysTyrProLeuGluAlaLysTyrGlnProLeuThrLeuProAspProTyrGlnLeuGlu 101
                 |||:::::::|||   :::   |||         |||   ::::::||||||
Db     333 AATGCAGAGCTGCAGTCAAAACTGAAGGATTTATTTAATGTGGATGCTTTTAAGCTGGAA 392

Qy     102 ----------------------------------AlaAlaPheIleLeuPheLys 108
                                               |||         :::
Db     393 TCATTAGAAGCAAAAAACAGAGCATTGAATGAACAGATTGCAAGATTGGAACAAGAAAGA 452

Qy     109 GluSerAspAlaAsnProAlaAsnSerThrGluLys---------------------- 120
           |||   :::   |||         |||      |||
Db     453 GAAAAAGAACCGAATCGTCTAGAGTCGTTGAGAAAACTGAAGGCTTCCTTACAAGGAGAT 512

Qy     121 ------ArgPheTrpMetArgPheArgArgGlyLysAsnHisSerTyrPheHisAspLeu 138
                :::::                 ::::::||||||        |||
Db     513 GTTCAAAAGTATCAGGCATACATGAGCAATTTGGAGTCTCATTCAGCCATTCTTGACCAG 572

Qy     139 ValPheAsnLeuLeuGluLysAsnValThrArg----------------------Asp 150
               |||   |||   :::   :::   |||
Db     573 AAATTAAATGGTCTCAATGAGGAAATTGCTAGAGTAGAACTAGAATGTGAAACAATAAAA 632

Qy     151 AlaAspAlaThrAspIleGluAsnPheAlaSerArgTyrLeuTyrMetAlaThrLeuTyr 170
              :::   |||   ::::::|||                    |||
Db     633 CAGGAGAACACTCGACTACAGAATATCATTGACAACCAGAAGTAC-------------- 677

Qy     171 TyrLysThrTyrThrAsnValAspGluPheGlyAlaSerPhePheAsnLysLeuSerPhe 190
               :::      :::::::::
Db     678 ------TCAGTTGCAGACATTGAGCGAATAAATCATGAA---------------------- 710

Qy     191 ThrThrGlyLeuPheGlyTrpGlyIleLysArgAlaLeuLysGlnIleIleArgSerAsn 210
                                        :::      |||:::|||   |||   :::
Db     711 --------------------------AGAAATGAATTGCAGCAGACTATT---AATAAA 740

Qy     211 LeuProLeuAspIleGlyThrGluHisSerValSerArgLeuGlnHisIleThrSerSer 230
           |||      |||:::      |||           :::|||   :::
Db     741 TTAACCAAGGACCTGGAAGCTGAACAA------CAGAAGTTGTGGAATGAGGAGTTAAAA 794

Qy     231 Tyr---------LysAspTyrMetAspThrGlnIleProAlaLeuProLysPheAlaLys 247
           |||         |||:::   ::::::||||||:::          |||   |||:::
Db     795 TATGCCAGAGGCAAAGAAGCGATTGAAACACAATTAGCAGAGTATCACAAATTGGCTAGA 854

Qy     248 ArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValAlaGlyTyrValAspThr 267
           :::      |||   ::::::   :::                  ||||||
Db     855 AAA------TTAAAACTTATTCCTAAAGGTGCTGAGAATTCCAAAGGTTAT---------- 899

Qy     268 ProTrpTyrLysLysTrpTyrMetLysLeuLys--------------AsnPheMetVal 282
                       :::   :::|||                     |||   :::|||
Db     900 -----------GACTTTGAAATTAAGTTTAATCCCGAGGCTGGTGCCAACTGCCTTGTC 947

Qy     283 Asn---------ArgValPheIleProThrLysLysPheAsnLysGluIleArgGlu 299
                  :::|||::::::|||   |||:::      |||:::         |||
Db     948 AAATACAGGGCTCAAGTTTATGTACCTCTTAAGGAACTCCTGAATGAAACTGAAGAAGAA 1007

Qy     300 ProSerLysAlaLeuLysGluLysValSer---ThrAspThrLysAsp------------ 314
              :::||||||||||   :::|||:::         ||||||   :::
Db    1008 ATTAATAAAGCCCTAAATAAAAAAATGGGTTTGGAGGATACTTTAGAACAATTGAATGCA 1067

Qy     315 ---LeuPheGluAsnLysIleGlyGlnGlyThrVal----------------------- 325
              :::   |||:::|||         ||||||:::
Db    1068 ATGATAACAGAAAGCAAGAGAAGTGTGGGAACTCTGAAAGAAGAAGTTCAAAAGCTGGAT 1127

Qy     326 AspPhePheAsnLysGluIleArgAspProSerLysAlaLeuLysGluLysValSerAsn 345
           |||   :::   ::::::|||::::::      :::         ||||||   ::::::
Db    1128 GATCTTTACCAACAAAAAATTAAGGAAGCAGAGGAAGAG---GATGAAAAATGTGCCAGT 1184

Qy     346 AspAla-----------------LysAspLeuPheGluAsnLysIleGlyGlnGlyThr 359
           :::                    |||   |||   |||:::   :::   ||||||
Db    1185 GAGCTTGAGTCCTTGGAGAAACACAAGCACCTGCTAGAAAGTACTGTTAACCAGGGGCTC 1244

Qy     360 ValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValSer 379
              :::   :::   ||||||:::         :::                |||
Db    1245 AGTGAAGCTATG---AATGAATTAGATGCTGTTCAGCGGGAATACCAACTAGTTGTGCAA 1301

Qy     380 ThrGlyAlaGluAspLeuPheGluAsnLysIleGly---------GlnGlyThrValAsp 396
           |||      |||:::            |||:::|||         |||      ::::::
Db    1302 ACCACGACTGAAGAA------AGACGAAAAGTGGGAAATAACTTGCAACGTCTGTTAGAG 1355

Qy     397 PheIleAsnAsnGluIleArgAspProSerLysAlaLeu---------IleArgLysVal 413
              :::       :::          |||   |||         |||   ||||||
Db    1356 ATGGTTGCTACACATGTTGGGTCTGTAGAGAAACATCTTGAGGAGCAGATTGCTAAAGTT 1415

Qy     414 ---------TyrThrGlu------AlaAspAspLeuPheGluAsnLysIleGlyGlnGly 428
```

```
                  |||   |||      ::::::||||||   ||||||
Db   1416 GATAGAGAATATGAAGAATGCATGTCAGAAGATCTCTCGGAAAAT-------------- 1460

Qy    429 ThrValAspPheIleAsnLysGluIleArgAsp------ProSerLysAlaLeuIleArg 446
                          ||||||||||||||||          ||||||   :::
Db   1461 --------------ATTAAAGAGATTAGAGATAAGTATGAGAAGAAAGCTACTCTAATT 1505

Qy    447 LysValSerThrGlu 451
          |||   |||   |||
Db   1506 AAGTCTTCTGAAGAA 1520

RESULT 30
AAT73285
ID   AAT73285 standard; cDNA; 3278 BP.
XX
AC   AAT73285;
XX
DT   12-SEP-1997  (first entry)
XX
DE   K. lactis origin of replication complex protein 1 gene.
XX
KW   Origin of replication complex; ORC; yeast; Kluyveromyces lactis;
KW   chromatography; peptide sequencing; primer; amplification; PCR; genome;
KW   polymerase chain reaction; open reading frame; cell growth; cancer;
KW   infection; inflammation; hypersensitivity; ds.
XX
OS   Kluyveromyces lactis.
XX
FH   Key             Location/Qualifiers
FT   CDS             395..3052
FT                   /*tag= a
FT                   /product= ORC 1 protein
XX
PN   US5614618-A.
XX
PD   25-MAR-1997.
XX
PF   16-DEC-1993;   93US-0168479.
XX
PR   07-JUN-1995;   95US-0484106.
PR   16-DEC-1993;   93US-0168479.
XX
PA   (COLD-) COLD SPRING HARBOR LAB.
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Bell SP,  Foss M,  Gavin K,  Herskowitz I,  Hidaka M;
PI   Kobayashi R,  Laurenson P,  Li J,  McNally FJ,  Rine J;
PI   Stillman BW;
XX
DR   WPI; 1997-201534/18.
DR   P-PSDB; AAW22230.
XX
PT   Nucleic acids encoding origin of replication complex proteins - used
PT   for screening for lead cpds. for therapy or diagnosis of disease
PT   associated with undesirable cell growth
XX
PS   Claim 4; Column 59-62; 53pp; English.
XX
CC   This is the nucleotide sequence encoding the origin of replication
CC   complex protein 1 (ORC1) from the yeast Kluyveromyces lactis.  The
CC   sequence was isolated using primers based on amino acid sequence
CC   conserved between the ORC1 and SIR3 proteins from Saccharomyces
CC   cerevisiae.  The amplified fragment was then used for low stringency
CC   DNA hybridisation to obtain the K. lactis ORC1 gene sequence.  The ORC
CC   proteins (AAW22224-35) can be used to screen chemical libraries to
CC   identify lead compounds useful in treatment and diagnosis of undesired
CC   cell growth, e.g. cancer, infections, inflammation and hypersensitivity.
XX
SQ   Sequence 3278 BP; 1085 A; 572 C; 714 G; 902 T; 5 other;

Alignment Scores:
Pred. No.:              0.705          Length:        3278
Score:                  111.00         Matches:       98
Percent Similarity:     35.50%         Conservative:  82
Best Local Similarity:  19.33%         Mismatches:    199
Query Match:            4.71%          Indels:        128
DB:                     18             Gaps:          22

US-09-807-459-2 (1-458) x AAT73285 (1-3278)

Qy      5 AspSerValGlyAspValThrLysThrLeuLeuAlaAlaSerGluSerValAspSerAla 24
          :::::::::   |||           |||           ||||||   :::
Db   1214 GAAGCTATATCAGATAATGAATCGGATTTATCTGAATATCACGAAAGTAAAGAAGAGTTT 1273

Qy     25 AlaAsnAlaTyrMetIleAsnSerAsp------MetSerAspTyrLeuSerAlaValSer 42
          |||||||||        :::||||||           ||||||   ||||||
Db   1274 GCAAACGCATCCTCTTCGGACAGTGATGAAGAGTTTGAAGATTACCAGTCTGCAGAAGAG 1333

Qy     43 AspAsnPheAlaGluArgIleCysSerGlnVal-------------------------- 53
                  |||         :::|||
Db   1334 CTTGCAATTGTAGAACCTGCCAAGAAAAAGGTGAGATCTATTAAACCAGATATACCCATT 1393

Qy     54 ---ProLysGlySerAsnCysSerAlaSerValSerAlaTyrMetSer---------Arg 69
             |||      |||                  ||||||      |||         :::
Db   1394 TCACCAGTAAAATCACAGACTCCATTGCAGCCATCAGCAGTTCATTCATCTCCTAGAAAG 1453

Qy     70 CysAlaLysGlnAspCysLeuThrLeuGlnSerLeuLysTyrProLeuGluAlaLysTyr 89
             |||   :::   :::      :::          |||         :::|||
Db   1454 TTCTTTAAGAATAATATAGTGCGCGCTAAAAAGGCATATACTCCATTTTCCAAACGGTAT 1513

Qy     90 GlnProLeuThrLeuProAspProTyrGlnLeuGluAlaAlaPheIleLeuPheLysGlu 109
          :::         :::||||||                    |||      :::
Db   1514 AAGAATCCGAAGATTCCTGACTTGAACGAT--------------ATTTTCCAAAGGCAT 1558

Qy    110 SerAspAlaAsnProAlaAsnSerThrGluLysArgPheTrpMetArgPheArgArgGly 129
          ::::::                 :::   |||:::||||||               :::|||
Db   1559 AATAATGATTTGGATATAGCTGCATTAGAGGAGAGATTCAGAACAGTTTCTGCTAAAGGC 1618

Qy    130 LysAsnHisSerTyrPheHisAspLeuValPheAsnLeuLeuGluLysAsnValThrArg 149
          |||      :::   |||      :::         |||      :::|||
Db   1619 AAAATGGAGACTATTTTTTCTAAGGTGAAGAAGCAATTGAACTCAAGGAATAGCAAAGAA 1678

Qy    150 Asp---AlaAspAlaThrAspIleGluAsnPhe---AlaSerArgTyrLeuTyrMetAla 167
          :::         |||   |||   :::|||:::     :::|||           |||
Db   1679 GAAATTGTCAAAGCTGCTGATTTCGACAATTATCTTCCGGCAAGAGAAAATGAATTTGCA 1738

Qy    168 ThrLeuTyrTyrLysThrTyrThrAsnValAspGluPheGlyAlaSerPhePheAsnLys 187
          :::::::|||         |||:::   :::   |||   |||   |||
Db   1739 AGTATATACCTCTCACTTTACAGTGCAATT---GAAGCAGGCACTAGCACCAGTATTTAC 1795

Qy    188 LeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAlaLeuLysGlnIleIle 207
          ::::::   |||   |||:::          |||           ::::::::::::::
Db   1796 ATTGCCGGGACGCCAGGCGTT---------GGTAAAACTTTGACGGTTCGAGAGGTAGTT 1846

Qy    208 ArgSerAsnLeuProLeuAspIleGlyThr---------GluHisSerValSerArgLeu 224
          :::              |||:::   |||         :::      :::   |||
Db   1847 AAG--------------GATTTAATGACATCTGCAGACCAAAAAGAACTTCCAAGATTC 1891

Qy    225 GlnHisIle-----------------------------ThrSerSerTyrLys------ 232
          |||:::|||                            :::   ||||||:::
Db   1892 CAATACATTGAAATCAATGGTTTAAAGATTGTCAAAGCAAGTGATAGTTATGAAGTCTTT 1951

Qy    233 ------------------------------------------AspTyrMetAspThr 237
                                                     :::   ::::::
Db   1952 TGGCAAAAAATATCTGGAGAAAAGCTTACATCTGGAGCTGCCATGGAATCTCTGGAGTTT 2011

Qy    238 GlnIleProAlaLeuProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeu 257
                :::|||       ||||||      ::::::::::::      |||
Db   2012 TATTTTAACAAAGTTCCAGCTACGAAAAAACGTCCTATCGTTGTGTTATTGGATGAGCTT 2071

Qy    258 LeuAlaThrValAlaGlyTyrValAspThrProTrp---TyrLysLysTrp------Tyr 274
            |||   |||:::          |||      :::   :::      |||      |||
Db   2072 GATGCATTAGTTAGCAAGAGCCAAGATGTAATGTACAACTTCTTTAACTGGGCTACCTAT 2131

Qy    275 MetLysLeuLysAsnPheMetVal---------AsnArgValPheIleProThrLysLys 291
                  |||      :::|||          |||   :::   :::|||   :::
```

```
Db   2132 TCAAATGCGAAACTTATTGTTGTAGCTGTCGCAAACACCTTAGATCTCCCCGAACGCCAT 2191

Qy    292 PhePheAsnLys---------------------------------------------- 295
          ||||||
Db   2192 CTTGGTAACAAGATTTCGTCCAGAATTGGTTTTACTAGAATTATGTTCACTGGTTACACG 2251

Qy    296 ------GluIleArg---------------------------------------------- 298
               |||:::|||
Db   2252 CATGAAGAGCTTAGAACAATCATCAATTTGAGACTTAAATATTTGAACGAATCTAGTTTC 2311

Qy    299 ------GluProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPhe 316
               :::|||                  :::|||   |||:::     :::
Db   2312 TATGTCGACCCGGAGACAGGGAGTTCGTACATGATCTCTCCGGATAGTAGTACTATA--- 2368

Qy    317 GluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspProSer 336
          |||          :::         ||||||   |||
Db   2369 GAAACTGATGAAGAAGAAAAGCGAAAAGACTTCTCTAAC------------------TAT 2410

Qy    337 LysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGly 356
          |||   ||||||   :::::::::   ||||||   ::::::         ||||||
Db   2411 AAACGACTAAAACTTAGGATTAATCCTGATGCCATTGAGATTGCATCAAGAAAAATTGCT 2470

Qy    357 GlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArg 376
                     ::::::   ::::::|||         |||   :::   |||
Db   2471 AGT--------------GTCAGTGGTGATGTGCGGAGAGCTTTAAAGGTGGTCAAAAGA 2515

Qy    377 LysValSerThrGlyAlaGluAspLeu--------------PheGluAsnLysIle--- 390
          |||              |||                  :::|||      :::
Db   2516 GCGGTAGAATATGCGGAAAATGATTACTTAAAGAGGCTTAGATATGAGCGACTAGTCAAT 2575

Qy    391 ----------------------GlyGlnGlyThrValAspPheIleAsnAsnGluIle 402
                               |||   |||      :::      :::   ||||||
Db   2576 TCCAAAAAAGATACTAGTGGCAATGGTACAGGAAATGAAGAATTACAGAGTGTAGAAATT 2635

Qy    403 ArgAspProSerLysAlaLeu 409
          :::      :::||||||||||
Db   2636 AAGCATATTACCAAGGCATTA 2656

RESULT 31
AAT62358
ID   AAT62358 standard; cDNA; 3278 BP.
XX
AC   AAT62358;
XX
DT   23-JUL-1997  (first entry)
XX
DE   Kluyveromyces lactis origin of replication complex ORC1 gene.
XX
KW   Origin of replication complex; ORC; gene therapy; cancer;
KW   neoplasia; inflammation; hypersensitivity; ds.
XX
OS   Kluyveromyces lactis.
XX
FH   Key             Location/Qualifiers
FT   misc_difference 309
FT                   /*tag=  a
FT                   /note= "base 309 is given as m in the
FT                          specification"
FT   misc_difference 311
FT                   /*tag=  b
FT                   /note= "base 311 is given as n in the
FT                          specification"
FT   misc_difference 327
FT                   /*tag=  c
FT                   /note= "base 327 is given as n in the
FT                          specification"
FT   misc_difference 328
FT                   /*tag=  d
FT                   /note= "base 328 is given as n in the
FT                          specification"
FT   misc_difference 3143
FT                   /*tag=  e
FT                   /note= "base 3143 is given as n in the
FT                          specification"
FT   CDS             395..1052
FT                   /*tag=  f
XX
PN   WO9640977-A1.
XX
PD   19-DEC-1996.
XX
PF   07-JUN-1996;   96WO-US09403.
XX
PR   07-JUN-1995;   95US-0484105.
XX
PA   (COLD-) COLD SPRING HARBOR LAB.
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Bell SP,  Foss M,  Herskowitz I,  Kobayashi R,  Laurenson P;
PI   Li JJ,  McNally FJ,  Rine J,  Stillman BW;
XX
DR   WPI; 1997-052354/05.
DR   P-PSDB; AAW14136.
XX
PT   Nucleic acid encoding origin of replication complex (ORC) protein -
PT   useful to screen for lead pharmaceuticals capable of disrupting ORC
PT   protein function, and inhibiting cell growth
XX
PS   Disclosure; Page 16-18; 57pp; English.
XX
CC   Isolated cDNA clones (AAT62358-63) respectively encode origin of
CC   replication (ORC) proteins (AAW14136-41) from Kluyveromyces lactis,
CC   Schizosaccharomyces pombe, human (ORC1), Arabidopsis thaliana,
CC   Caenorhabditis elegans and human (ORC2).  The K. lactis ORC1 clone
CC   was obtd. by PCR amplification using primers based on Saccharomyces
CC   cerevisiae ORC1 and SIR3 sequences.  The isolated nucleic acids can
CC   be utilised in the prodn. of ORC polypeptides, to design probes and
CC   primers for the detection and amplification of ORC genes, and in
CC   gene therapy applns., e.g. antisense oligonucleotides capable of
CC   inhibiting the intracellular expression of a targetted ORC
CC   transcript.
XX
SQ   Sequence 3278 BP; 1085 A; 572 C; 714 G; 902 T; 5 other;

Alignment Scores:
Pred. No.:              0.705          Length:         3278
Score:                  111.00         Matches:        98
Percent Similarity:     35.50%         Conservative:   82
Best Local Similarity:  19.33%         Mismatches:     199
Query Match:            4.71%          Indels:         128
DB:                     18             Gaps:           22

US-09-807-459-2 (1-458) x AAT62358 (1-3278)

Qy      5 AspSerValGlyAspValThrLysThrLeuLeuAlaAlaSerGluSerValAspSerAla 24
          :::::::::   |||            |||            ||||||   :::
Db   1214 GAAGCTATATCAGATAATGAATCGGATTTATCTGAATATCACGAAAGTAAAGAAGAGTTT 1273

Qy     25 AlaAsnAlaTyrMetIleAsnSerAsp------MetSerAspTyrLeuSerAlaValSer 42
          |||||||||         :::||||||           ||||||   ||||||
Db   1274 GCAAACGCATCCTCTTCGGACAGTGATGAAGAGTTTGAAGATTACCAGTCTGCAGAAGAG 1333

Qy     43 AspAsnPheAlaGluArgIleCysSerGlnVal--------------------------- 53
                 |||           :::|||
Db   1334 CTTGCAATTGTAGAACCTGCCAAGAAAAAGGTGAGATCTATTAAACCAGATATACCCATT 1393

Qy     54 ---ProLysGlySerAsnCysSerAlaSerValSerAlaTyrMetSer---------Arg 69
             |||      |||                |||||      |||         :::
Db   1394 TCACCAGTAAAATCACAGACTCCATTGCAGCCATCAGCAGTTCATTCATCTCCTAGAAAG 1453

Qy     70 CysAlaLysGlnAspCysLeuThrLeuGlnSerLeuLysTyrProLeuGluAlaLysTyr 89
                |||   :::   :::       :::          |||         :::|||
Db   1454 TTCTTTAAGAATAATATAGTGCGCGCTAAAAAGGCATATACTCCATTTTCCAAACGGTAT 1513

Qy     90 GlnProLeuThrLeuProAspProTyrGlnLeuGluAlaAlaPheIleLeuPheLysGlu 109
          :::         :::||||||                          |||      :::
```

```
Db   1514 AAGAATCCGAAGATTCCTGACTTGAACGAT--------------ATTTTCCAAAGGCAT 1558

Qy    110 SerAspAlaAsnProAlaAsnSerThrGluLysArgPheTrpMetArgPheArgArgGly 129
          ::::::                 :::   ||||::|||||                ::|||
Db   1559 AATAATGATTTGGATATAGCTGCATTAGAGGAGAGATTCAGAACAGTTTCTGCTAAAGGC 1618

Qy    130 LysAsnHisSerTyrPheHisAspLeuValPheAsnLeuLeuGluLysAsnValThrArg 149
          |||      :::   |||      :::         |||      :::|||
Db   1619 AAAATGGAGACTATTTTTTCTAAGGTGAAGAAGCAATTGAACTCAAGGAATAGCAAAGAA 1678

Qy    150 Asp---AlaAspAlaThrAspIleGluAsnPhe---AlaSerArgTyrLeuTyrMetAla 167
          :::         |||   |||   :::|||:::      :::|||            |||
Db   1679 GAAATTGTCAAAGCTGCTGATTTCGACAATTATCTTCCGGCAAGAGAAAATGAATTTGCA 1738

Qy    168 ThrLeuTyrTyrLysThrTyrThrAsnValAspGluPheGlyAlaSerPhePheAsnLys 187
          ::::::|||       |||:::  :::  |||  |||  |||
Db   1739 AGTATATACCTCTCACTTTACAGTGCAATT---GAAGCAGGCACTAGCACCAGTATTTAC 1795

Qy    188 LeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAlaLeuLysGlnIleIle 207
          ::::::  |||  |||:::         |||           ::::::::::::::
Db   1796 ATTGCCGGGACGCCAGGCGTT---------GGTAAAACTTTGACGGTTCGAGAGGTAGTT 1846

Qy    208 ArgSerAsnLeuProLeuAspIleGlyThr---------GluHisSerValSerArgLeu 224
          :::               |||:::  |||         :::      :::   |||
Db   1847 AAG--------------GATTTAATGACATCTGCAGACCAAAAAGAACTTCCAAGATTC 1891

Qy    225 GlnHisIle-----------------------------ThrSerSerTyrLys------ 232
          |||:::|||                            :::   ||||||:::
Db   1892 CAATACATTGAAATCAATGGTTTAAAGATTGTCAAAGCAAGTGATAGTTATGAAGTCTTT 1951

Qy    233 ------------------------------------------AspTyrMetAspThr 237
                                                     :::   ::::::
Db   1952 TGGCAAAAAATATCTGGAGAAAAGCTTACATCTGGAGCTGCCATGGAATCTCTGGAGTTT 2011

Qy    238 GlnIleProAlaLeuProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeu 257
                 :::|||         ||||||      ::::::::::::      |||
Db   2012 TATTTTAACAAAGTTCCAGCTACGAAAAAACGTCCTATCGTTGTGTTATTGGATGAGCTT 2071

Qy    258 LeuAlaThrValAlaGlyTyrValAspThrProTrp---TyrLysLysTrp------Tyr 274
             |||   |||:::         |||      :::   :::      |||      |||
Db   2072 GATGCATTAGTTAGCAAGAGCCAAGATGTAATGTACAACTTCTTTAACTGGGCTACCTAT 2131

Qy    275 MetLysLeuLysAsnPheMetVal---------AsnArgValPheIleProThrLysLys 291
                |||      :::|||         |||  :::   :::|||   :::
Db   2132 TCAAATGCGAAACTTATTGTTGTAGCTGTCGCAAACACCTTAGATCTCCCCGAACGCCAT 2191

Qy    292 PhePheAsnLys---------------------------------------------- 295
                |||||
Db   2192 CTTGGTAACAAGATTTCGTCCAGAATTGGTTTTACTAGAATTATGTTCACTGGTTACACG 2251

Qy    296 ------GluIleArg------------------------------------------- 298
                |||:::|||
Db   2252 CATGAAGAGCTTAGAACAATCATCAATTTGAGACTTAAATATTTGAACGAATCTAGTTTC 2311

Qy    299 ------GluProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPhe 316
                :::|||                    :::|||   |||:::      :::
Db   2312 TATGTCGACCCGGAGACAGGGAGTTCGTACATGATCTCTCCGGATAGTAGTACTATA--- 2368

Qy    317 GluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspProSer 336
          |||             :::         ||||||   |||
Db   2369 GAAACTGATGAAGAAGAAAAGCGAAAAGACTTCTCTAAC-----------------TAT 2410

Qy    337 LysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGly 356
          |||   ||||||   :::::::::   ||||||   ::::::        ||||||
Db   2411 AAACGACTAAAACTTAGGATTAATCCTGATGCCATTGAGATTGCATCAAGAAAAATTGCT 2470

Qy    357 GlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArg 376
                      ::::::   ::::::|||         |||   :::   |||
Db   2471 AGT--------------GTCAGTGGTGATGTGCGGAGAGCTTTAAAGGTGGTCAAAAGA 2515

Qy    377 LysValSerThrGlyAlaGluAspLeu--------------PheGluAsnLysIle--- 390
             |||               |||                ::::|||      :::
Db   2516 GCGGTAGAATATGCGGAAAATGATTACTTAAAGAGGCTTAGATATGAGCGACTAGTCAAT 2575
```

```
Qy    391 ----------------------GlyGlnGlyThrValAspPheIleAsnAsnGluIle 402
                                 |||   |||      :::      :::   ||||||
Db   2576 TCCAAAAAAGATACTAGTGGCAATGGTACAGGAAATGAAGAATTACAGAGTGTAGAAATT 2635

Qy    403 ArgAspProSerLysAlaLeu 409
          :::      :::||||||||||
Db   2636 AAGCATATTACCAAGGCATTA 2656

RESULT 32
AAQ06842
ID   AAQ06842 standard; RNA; 1777 BP.
XX
AC   AAQ06842;
XX
DT   05-MAR-1991  (first entry)
XX
DE   RNA segment 4 coding for HA.
XX
KW   Influenza; haemagglutinin; HA; ribosomal frameshift signal sequence;
KW   membrane anchor; RFS; ss.
XX
OS   Influenza virus A/PR8/34.
XX
FH   Key             Location/Qualifiers
FT   CDS             33..1732
FT                   /*tag= a
FT   sig_peptide     33..82
FT                   /*tag= b
FT   mat_peptide     83..1063
FT                   /*tag= c
FT                   /product= HA1
FT   mat_peptide     1064..1730
FT                   /*tag= d
FT                   /product= HA2
FT   misc_RNA        1632..1692
FT                   /*tag= e
FT                   /label= insertion site for RFS
XX
PN   WO9014422-A.
XX
PD   29-NOV-1990.
XX
PF   21-MAY-1990;   90WO-GB00791.
XX
PR   19-MAY-1989;   89GB-0011555.
XX
PA   (LYNX-) LYNXVALE LTD.
XX
PI   Inglis SC,  Brierley I;
XX
DR   WPI; 1990-375989/50.
XX
PT   Ribosomal frame shifting signal sequences - isolated from
PT   infectious bronchitis virus genomic RNA and used in protein
PT   prodn.
XX
PS   Disclosure; Fig 19; 55pp; English.
XX
CC   The HA gene encodes a spike-like protein which is embedded in the
CC   membrane via a hydrophobic anchor sequence.  A portion of this
CC   anchor sequence may be replaced with a ribosomal frame shift signal
CC   sequence (RFS), in such a way that ribosomes translating the new
CC   HA sequence will usually terminate before the hydrophobic sequence
CC   is encountered, leading to the prodn. of a secreted form of the HA.
CC   It has been found that the primary sequence of the FRS can be rad-
CC   ically altered as long as the the secondary and tertiary structures
CC   are preserved, so it is possible to design an RFS which encodes
CC   hydrophobic amino acids, and therefore preserves the integrity of
CC   the anchor.
CC   See also AAQ06841 and AAQ07007.
XX
SQ   Sequence 1777 BP; 621 A; 331 C; 408 G; 417 U; 0 other;
```

```
Alignment Scores:
Pred. No.:              0.391          Length:         1777
Score:                  110.00         Matches:        97
Percent Similarity:     35.04%         Conservative:   67
Best Local Similarity:  20.73%         Mismatches:     170
Query Match:            4.66%          Indels:         134
DB:                     11             Gaps:           24

US-09-807-459-2 (1-458) x AAQ06842 (1-1777)

Qy      23 SerAlaAlaAsnAlaTyrMetIleAsnSerAspMetSerAspTyr---------------- 37
           ::::::   |||                  |||      ||||||
Db     332 AACUCUGAGAAUGGAAUAUGUUAUCCAGGAGAUUUCAUCGACUAUGAGGAGCUGAGGGAG 391

Qy      38 ---LeuSerAlaValSerAspAsnPheAlaGluArgIleCysSerGlnValProLysGly 56
              |||||||:::||||||          ||||||              ||||||
Db     392 CAAUUGAGCUCAGUGUCAUCA------UUCGAAAGAUUC---GAAAUAUUUCCCAAAGAA 442

Qy      57 Ser---------------AsnCysSerAlaSerValSerAlaTyrMetSerArgCysAla 71
           |||               |||   :::      |||:::|||      |||
Db     443 AGCUCAUGGCCCAACCACAACACAACCAAAGGAGUAACGGCAGCAUGCUCCCAUGCGGGG 502

Qy      72 LysGlnAsp-----------------CysLeuThrLeuGlnSerLeuLysTyrProLeu 85
           |||                      ||||||   :::      ||||||
Db     503 AAAAGCAGUUUUUACAGAAAUUUGCUAUGGCUGACGGAGAAGGAGGGCUCAUACCCA--- 559

Qy      86 GluAlaLysTyrGlnProLeuThrLeuProAspProTyrGlnLeuGluAlaAlaPheIle 105
                                                 :::|||:::   :::::::::
Db     560 ------------------------------------AAGCUGAAAAAUUCUUAUGUG 580

Qy     106 LeuPheLysGluSerAspAla--------------------AsnProAlaAsnSerThr 118
                ||||      :::                      :::|||:::||||||
Db     581 AACAAGAAAGGGAAAGAAGUCCUUGUACUGUGGGGUAUUCAUCACCCGUCUAACAGUAAG 640

Qy     119 GluLysArgPheTrpMetArgPheArgArgGlyLysAsnHisSerTyrPheHisAspLeu 138
           :::::::::            :::   :::|||   :::|||       :::
Db     641 GAUCAACAG---------AAUAUCUAUCAGAAUGAAAAU---GCUUAUGUCUCUGUAGUG 688

Qy     139 ValPheAsnLeuLeuGluLysAsnValThrArgAspAlaAspAlaThrAspIleGluAsn 158
                |||       :::              |||:::         :::   :::
Db     689 ACUUCAAAUUAUAACAGGAGAUUUACCCCGGAAAUAGCAGAAAGACCCAAAGUAAGAGAU 748

Qy     159 PheAlaSerArgTyrLeuTyrMetAlaThrLeuTyrTyrLysThrTyrThrAsnValAsp 178
              |||   |||      |||      ||||||               |||   :::
Db     749 CAAGCUGGGAGGAUGAACUAUUACUGGACCUUGCUAAAACCCGGAGACACAAUAAUAUUU 808

Qy     179 GluPheGlyAlaSerPhePhe---------AsnLysLeuSerPheThrThrGlyLeuPhe 195
           |||       :::                  :::   :::   |||   |||
Db     809 GAGGCAAAUGGAAAUCUAAUAGCACCAAGGUAUGCUUUCGCACUGAGUAGAGGC---UUU 865

Qy     196 GlyTrpGlyIleLysArgAlaLeuLysGlnIleIleArgSerAsnLeuProLeuAspIle 215
           |||   |||                  ||||||   ||||||      :::
Db     866 GGGUCCGGC-------------------AUCAUCACCUCAAACGCAUCAAUGCAUGAG 904

Qy     216 GlyThrGluHisSerValSerArgLeuGlnHisIleThrSerSer------TyrLysAsp 233
                       :::   |||      |||   ||||||      :::::::::
Db     905 UGUAACACGAAGUGUCAAACACCCCUGGGAGCUAUAAACAGCAGUCUCCCUUUCCAGAAU 964

Qy     234 TyrMetAspThrGlnIleProAlaLeuProLysPheAlaLysArgPheSerLeuMetVal 253
                  |||         ||||||:::   :::         |||   :::
Db     965 AUACACCCAGUCACAAUAGGAGAGUGCCCAAAAUACGUCAGGAGUGCCAAAUUGAGGAUG 1024

Qy     254 Val--------------------------------GlnArgLeuLeuAlaThrValAla 262
           |||                              :::   |||         :::|||
Db    1025 GUUACAGGACUAAGGAACAUUCCGUCCAUUCAAUCCAGAGGUCUAUUUGGAGCCAUUGCC 1084

Qy     263 GlyTyrValAspThrProTrp---------TyrLysLysTrpTyr--------------- 274
           |||:::::::::      |||              ||||||
Db    1085 GGUUUUAUUGAAGGGGGAUGGACUGGAAUGAUAGAUGGAUGGUACGGUUAUCAUCAUCAG 1144

Qy     275 ------------------------------------------MetLysLeuLysAsnPhe 280
                                                       |||
```

```
Db    1145 AAUGAACAGGGAUCAGGCUAUGCAGCGGAUCAAAAAAGCACACAAAAUGCCAUUAACGGG 1204

Qy     281 MetValAsnArgVal----------------------------PheIleProThrLys 290
           :::   |||:::|||                           |||
Db    1205 AUUACAAACAAGGUGAACUCUGUUAUCGAGAAAAUGAACAUUCAAUUCACAGCUGUGGGU 1264

Qy     291 LysPhePheAsnLysGluIleArgGluProSerLysAlaLeuLysGluLysValSerThr 310
           |||   |||||||||   :::   :::      :::   |||   :::|||||||
Db    1265 AAAGAAUUCAACAAA---UUAGAAAAAAGGAUGGAAAAUUUAAAUAAAAAAGUUGAUGAU 1321

Qy     311 AspThrLysAsp----------------------------LeuPheGluAsnLysIle 320
                |||                               |||   ||||||
Db    1322 GGAUUUCUGGACAUUUGGACAUAUAAUGCAGAAUUGUUAGUUCUACUGGAAAAU------ 1375

Qy     321 GlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspProSerLysAlaLeuLys 340
              :::   |||:::||||||   :::      :::::::::      :::   :::|||
Db    1376 ---GAAAGGACUCUGGAUUUCCAUGACUCAAAUGUGAAGAAUCUGUAUGAGAAAGUAAAA 1432

Qy     341 GluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThrVal 360
              ::::::   |||:::||||||:::              ||||||   |||
Db    1433 AGCCAAUUAAAGAAUAAUGCCAAAGAA--------------AUCGGAAAUGGAUGUUUU 1477

Qy     361 AspPhe---------IleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLys 377
           :::|||           :::||||||                         :::
Db    1478 GAGUUCUACCACAAGUGUGACAAUGAA--------------------UGCAUGGAAAGU 1516

Qy     378 ValSerThrGlyAlaGluAspLeu-----------PheGluAsnLysIleGlyGlnGly 393
           |||      |||      |||                |||:::|||:::   :::
Db    1517 GUAAGAAAUGGGACUUAUGAUUAUCCCAAAUAUUCAGAAGAGUCAAAGUUGAACAGGGAA 1576

Qy     394 ThrValAspPheIleAsnAsnGlu 401
              ||||||   :::      |||
Db    1577 AAGGUAGAUGGAGUGAAAUUGGAA 1600

RESULT 33
AAS74418
ID   AAS74418 standard; cDNA; 1816 BP.
XX
AC   AAS74418;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #10222.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG10231.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 10222; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
```

```
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1816 BP; 726 A; 380 C; 328 G; 382 T; 0 other;

Alignment Scores:
Pred. No.:              0.45          Length:          1816
Score:                  109.50        Matches:         76
Percent Similarity:     37.99%        Conservative:    41
Best Local Similarity:  24.68%        Mismatches:      113
Query Match:            4.64%         Indels:          78
DB:                     23            Gaps:            16

US-09-807-459-2 (1-458) x AAS74418 (1-1816)

Qy     183 SerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAla 202
           :::||||||   :::|||     ||||||      |||   |||      |||||||||
Db     721 ACTTTCTTCACAGAATTGGAAAAAACTACTTTAAAGTTCAAATGGAACCAAAAAAGAGCC 780

Qy     203 LeuLysGlnIleIleArgSerAsnLeu--ProLeuAspIleGlyThrGluHisSerValS 222
                 |||   :::|||   |||  |||   :::          |||      :
Db     781 TGC------ATTGCCAAGTCAATCCTAAGCCAAAAGAACAAAGCTGGAGGCATCATGCTA 834

Qy     222 erArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaL 242
           ::|||||||||   |||         |||   :::              :::      :
Db     835 CCCGACTTCAAACTATACTACAAGGCTACAGTAAT------CAAAACAGCATGGTACTGG 888

Qy     242 euProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValA 262
           ::||||||      :::|||:::      |||          |||   |||          |
Db     889 TACCAAAA------CAGAGATATAGACCAATGGAACAGAACAGAGCCCTCAGAAATAATG 942

Qy     262 laGlyTyr---------------Val-AspThrPro---------------Trp------ 269
           ||   |||              ::: |||   |||               |||
Db     943 CCGCATATCTACAACTATCTGATCTTCGACAAACCTGACCAAAACAAGCAATGGGGAAAG 1002

Qy     270 ---------TyrLysLysTrpTyr-----------------------------MetLys 276
                   :::   |||||||||                             :::|||
Db     1003 GATTCCCTATTTAATAAATGGTACTGGGAAAACTGCCTAGCCATATGTAGAAAGCTGAAA 1062

Qy     277 LeuLysAsnPheMetValAsnArgValPheIleProThrLysLysPhePheAsnLysGlu 296
           |||      |||:::                |||   ::::::       :::
Db     1063 CTAGATCCCTTCCTTACACCTTATACAAAAATTAATTCAAGATGGATTAAAGACTTAAAT 1122

Qy     297 IleArg---GluProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeu 315
           :::|||   :::       |||||||||:::|||   :::           :::|||
Db     1123 GTTAGACTTAAAACCATAAAAGCCCTAGAAGAAAACCTAGGCAATACCATTCAGGAC--- 1179

Qy     316 PheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspPro 335
                   ||||||   |||      ||||||   :::|||
Db     1180 -----------ATAGGCATGGGCAAG---GACTTCATGTCTAAA-----------ACA 1212

Qy     336 SerLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLys--- 354
             ||||||:::      |||         |||      ||||||   :::   |||
Db     1213 CCAAAAGCAATGGCAACAAAAGACAAAATTGACAAATGGGATCTAATTAAACCAAAGAGC 1272

Qy     355 ------IleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLys 372
                    :::   |||         :::|||         |||   |||:::|||
Db     1273 TTCTGCACAGCAAAAGAAACTACCATCAGAGTGAAC---------AGGCAACCTACAAAA 1323

Qy     373 AlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGln 392
                                        |||   :::|||
Db     1324 --------------------------TGGGAGAAAATTTTCGCAACC------------ 1344

Qy     393 GlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLys 412
                                                  |||   ||||||   :::
Db     1345 ----------------------------TACTCATCTGACAAAGGGCTAATATCCAGA 1374

Qy     413 ValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhe 432
               |||   |||         :::::::::   |||            |||:::
Db     1375 AACTACAATGAACTCAAACAAATTTACAAGAAAAAA-----------ACAATCGGCCCC 1422

Qy     433 IleAsnLysGluIleArgAspProSerLysAlaLeuIleArgLysValSerThrGluAla 452
           |||   |||         |||   ::::::         ::::::   :::      |||
Db     1423 ATCAAAAAGTGGGCAGAGGATATGAACAGGCACTTCTCAAAAGAAGACACTTATGCAGCC 1482

Qy     453 AspAsnLeuLeuGluLys 458
                   :::::::|||
Db     1483 AAAAAACACATGAAAAAA 1500

RESULT 34
AAS69641
ID   AAS69641 standard; cDNA; 2277 BP.
XX
AC   AAS69641;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #5445.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG05454.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 5445; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
```

```
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2277 BP; 878 A; 483 C; 473 G; 443 T; 0 other;

Alignment Scores:
Pred. No.:              0.609          Length:          2277
Score:                  109.50         Matches:         73
Percent Similarity:     37.21%         Conservative:    39
Best Local Similarity:  24.25%         Mismatches:      107
Query Match:            4.64%          Indels:          82
DB:                     23             Gaps:            15

US-09-807-459-2 (1-458) x AAS69641 (1-2277)

Qy    183 SerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAla 202
          :::|||||   :::|||      ||||||      |||   |||      |||||||||
Db    868 ACTTTCTTCCCAGAATTGGAAAAAACTACTTTAAAGTTCATATGGAACCAAAAAAGAGCC 927

Qy    203 LeuLysGlnIleIleArgSerAsnLeu--ProLeuAspIleGlyThrGluHisSerValS 222
          |||      |||   :::|||   |||  |||   :::         :::|||      :
Db    928 CTC------ATCGCCAAGTCAATCCTGAGCCAAAAGAACAAAGCCAGAAGCATCACGCTA 981

Qy    222 erArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaL 242
          ::   |||:::   |||         |||   :::                 :::      :
Db    982 CCTGACTTCGAACTATACTACAAGGCTACAGTAAC------CAAAACAGCCTGGTACTGG 1035

Qy    242 euProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValA 262
          ::||||||    :::|||:::      |||         |||   |||      |
Db    1036 TACCAAAA------CAGAGATATAGACCAATGGAACAGAACAGAGCCCTCAGAAATAATG 1089

Qy    262 laGlyTyr---------------Val-AspThrPro------------------------ 268
          ||   |||               ::: |||   |||
Db    1090 CCGCATATCTACAACTATCTGATCTTTGACAAACCTCACAAAAACAAGCAATGGGGAAAG 1149

Qy    269 ------------------Trp---TyrLysLysTrpTyrMetLysLeuLysAsnPheMet 281
                           |||   ::::::||||||            :::
Db    1150 GATTCCCTATTTAATAAATGGTGCTGGGAAAAATGGCTAACCATATGTAGAAAGCTGAAA 1209

Qy    282 ValAsnArgValPheIleProThrLysLysPhePheAsnLysGluIleArgGlu------ 299
          ::::::              |||      |||      ::::::   |||::::::
Db    1210 CTGGATCCCTTCCTTACACCTTATACAAAAATTAATTCAAGATGGATTAAAGACTTACAT 1269

Qy    300 ------------ProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeu 315
                       |||   |||:::||||||:::            :::|||
Db    1270 GTTAGAGCTAAAACCATAAAAACCCTAGAAGAAAAACTAGGCAATACCATTCAGGAC--- 1326

Qy    316 PheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspPro 335
                |||||   |||      ||||||   :::|||            |||
Db    1327 ------------ATAGGCATGGGCAAG---GACTTCATGTCTAAA-----------CCA 1359

Qy    336 SerLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLys--- 354
          ||||||         |||      |||      ||||||   :::   |||
Db    1360 CCAAAAGCAACGGCAGCAAAAGCCAAAATTGACAAATGGGATCTAATTAAACTAAAGAAC 1419

Qy    355 ------IleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLys 372
                :::   |||         :::|||         |||   |||:::|||
Db    1420 TTCTGCACAGCAAAAGAAACTACCATCAGAGTGAAC---------AGGCAACCTACAAAA 1470

Qy    373 AlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGln 392
                                     |||   :::|||
Db    1471 -------------------------TGGGAGAAAATTTTCACAACC------------ 1491

Qy    393 GlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLys 412
                                              |||   ||||||   :::
```

```
Db    1492 -------------------------------TACTCATCTGACAAAGGGCTAATATCCAGA 1521

Qy     413 ValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhe 432
           :::|||   |||         :::::::::   |||
Db    1522 ATCTACAATGAACTCCAGCAAATTTACAAGAAAAAA---------------------- 1557

Qy     433 IleAsnLysGluIleArgAspProSerLysAlaLeuIleArgLysValSerThrGlu 451
             |||      |||:::      :::|||   :::   |||      |||   |||
Db    1558 ACAAACAACCCCATCAAAAAGTGGGCGAAGGACATGAACAGACACTTCTCAAAAGAA 1614

RESULT 35
AAS71279
ID   AAS71279 standard; cDNA; 2277 BP.
XX
AC   AAS71279;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #7083.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG07092.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 7083; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2277 BP; 878 A; 483 C; 473 G; 443 T; 0 other;

Alignment Scores:
```

```
Pred. No.:              0.609          Length:          2277
Score:                  109.50         Matches:         73
Percent Similarity:     37.21%         Conservative:    39
Best Local Similarity:  24.25%         Mismatches:      107
Query Match:            4.64%          Indels:          82
DB:                     23             Gaps:            15

US-09-807-459-2 (1-458) x AAS71279 (1-2277)

Qy    183 SerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAla 202
          :::||||||   :::|||      ||||||      |||   |||      |||||||||
Db    868 ACTTTCTTCCCAGAATTGGAAAAAACTACTTTAAAGTTCATATGGAACCAAAAAAGAGCC 927

Qy    203 LeuLysGlnIleIleArgSerAsnLeu--ProLeuAspIleGlyThrGluHisSerValS 222
          |||      |||   :::|||   |||  |||   :::         :::|||      :
Db    928 CTC------ATCGCCAAGTCAATCCTGAGCCAAAAGAACAAAGCCAGAAGCATCACGCTA 981

Qy    222 erArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaL 242
          ::   |||:::   |||        |||  :::                 :::      :
Db    982 CCTGACTTCGAACTATACTACAAGGCTACAGTAAC------CAAAACAGCCTGGTACTGG 1035

Qy    242 euProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValA 262
          ::||||||     :::|||:::      |||        |||   |||         |
Db   1036 TACCAAAA------CAGAGATATAGACCAATGGAACAGAACAGAGCCCTCAGAAATAATG 1089

Qy    262 laGlyTyr--------------Val-AspThrPro----------------------- 268
          ||   |||              ::: |||   |||
Db   1090 CCGCATATCTACAACTATCTGATCTTTGACAAACCTCACAAAAACAAGCAATGGGGAAAG 1149

Qy    269 -----------------Trp---TyrLysLysTrpTyrMetLysLeuLysAsnPheMet 281
                          |||   ::::::||||||            :::
Db   1150 GATTCCCTATTTAATAAATGGTGCTGGGAAAAATGGCTAACCATATGTAGAAAGCTGAAA 1209

Qy    282 ValAsnArgValPheIleProThrLysLysPheAsnLysGluIleArgGlu------ 299
          ::::::           |||      |||      ::::::   |||::::::
Db   1210 CTGGATCCCTTCCTTACACCTTATACAAAAATTAATTCAAGATGGATTAAAGACTTACAT 1269

Qy    300 ------------ProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeu 315
                       |||   |||:::||||||:::            :::|||
Db   1270 GTTAGAGCTAAAACCATAAAAACCCTAGAAGAAAAACTAGGCAATACCATTCAGGAC--- 1326

Qy    316 PheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspPro 335
          ||||||   |||      ||||||   :::|||            |||
Db   1327 ------------ATAGGCATGGGCAAG---GACTTCATGTCTAAA------------CCA 1359

Qy    336 SerLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLys--- 354
             ||||||         |||        |||      ||||||   :::   |||
Db   1360 CCAAAAGCAACGGCAGCAAAAGCCAAAATTGACAAATGGGATCTAATTAAACTAAAGAAC 1419

Qy    355 ------IleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLys 372
                ::::   |||         :::|||         |||   |||:::|||
Db   1420 TTCTGCACAGCAAAAGAAACTACCATCAGAGTGAAC---------AGGCAACCTACAAAA 1470

Qy    373 AlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGln 392
                                      |||   :::|||
Db   1471 --------------------------TGGGAGAAAATTTTCACAACC------------ 1491

Qy    393 GlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLys 412
                                                 |||   ||||||   :::
Db   1492 ----------------------------TACTCATCTGACAAAGGGCTAATATCCAGA 1521

Qy    413 ValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhe 432
          :::|||   |||         :::::::::   |||
Db   1522 ATCTACAATGAACTCCAGCAAATTTACAAGAAAAAA---------------------- 1557

Qy    433 IleAsnLysGluIleArgAspProSerLysAlaLeuIleArgLysValSerThrGlu 451
             |||      |||:::      :::|||   :::   |||      |||   |||
Db   1558 ACAAACAACCCCATCAAAAAGTGGGCGAAGGACATGAACAGACACTTCTCAAAAGAA 1614

RESULT 36
AAS74306
ID   AAS74306 standard; cDNA; 2277 BP.
XX
AC   AAS74306;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #10110.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG10119.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 10110; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2277 BP; 878 A; 483 C; 473 G; 443 T; 0 other;

Alignment Scores:
Pred. No.:              0.609          Length:          2277
Score:                  109.50         Matches:         73
Percent Similarity:     37.21%         Conservative:    39
Best Local Similarity:  24.25%         Mismatches:      107
Query Match:            4.64%          Indels:          82
DB:                     23             Gaps:            15

US-09-807-459-2 (1-458) x AAS74306 (1-2277)

Qy    183 SerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAla 202
          :::||||||   :::|||      ||||||      |||   |||      |||||||||
Db    868 ACTTTCTTCCCAGAATTGGAAAAAACTACTTTAAAGTTCATATGGAACCAAAAAAGAGCC 927

Qy    203 LeuLysGlnIleIleArgSerAsnLeu--ProLeuAspIleGlyThrGluHisSerValS 222
```

```
Db    928 CTC------ATCGCCAAGTCAATCCTGAGCCAAAAGAACAAAGCCAGAAGCATCACGCTA 981

Qy    222 erArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaL 242
Db    982 CCTGACTTCGAACTATACTACAAGGCTACAGTAAC------CAAAACAGCCTGGTACTGG 1035

Qy    242 euProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValA 262
Db   1036 TACCAAAA------CAGAGATATAGACCAATGGAACAGAACAGAGCCCTCAGAAATAATG 1089

Qy    262 laGlyTyr---------------Val-AspThrPro------------------------ 268
Db   1090 CCGCATATCTACAACTATCTGATCTTTGACAAACCTCACAAAAACAAGCAATGGGGAAAG 1149

Qy    269 ------------------Trp---TyrLysLysTrpTyrMetLysLeuLysAsnPheMet 281
Db   1150 GATTCCCTATTTAATAAATGGTGCTGGGAAAAATGGCTAACCATATGTAGAAAGCTGAAA 1209

Qy    282 ValAsnArgValPheIleProThrLysLysPhePheAsnLysGluIleArgGlu------ 299
Db   1210 CTGGATCCCTTCCTTACACCTTATACAAAAATTAATTCAAGATGGATTAAAGACTTACAT 1269

Qy    300 ------------ProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeu 315
Db   1270 GTTAGAGCTAAAACCATAAAAACCCTAGAAGAAAAACTAGGCAATACCATTCAGGAC--- 1326

Qy    316 PheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspPro 335
Db   1327 ------------ATAGGCATGGGCAAG---GACTTCATGTCTAAA-----------CCA 1359

Qy    336 SerLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLys--- 354
Db   1360 CCAAAAGCAACGGCAGCAAAAGCCAAAATTGACAAATGGGATCTAATTAAACTAAAGAAC 1419

Qy    355 ------IleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLys 372
Db   1420 TTCTGCACAGCAAAAGAAACTACCATCAGAGTGAAC---------AGGCAACCTACAAAA 1470

Qy    373 AlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGln 392
Db   1471 --------------------------TGGGAGAAAATTTTCACAACC------------ 1491

Qy    393 GlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLys 412
Db   1492 ----------------------------TACTCATCTGACAAAGGGCTAATATCCAGA 1521

Qy    413 ValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhe 432
Db   1522 ATCTACAATGAACTCCAGCAAATTTACAAGAAAAAA---------------------- 1557

Qy    433 IleAsnLysGluIleArgAspProSerLysAlaLeuIleArgLysValSerThrGlu 451
Db   1558 ACAAACAACCCCATCAAAAAGTGGGCGAAGGACATGAACAGACACTTCTCAAAAGAA 1614
```

```
RESULT 37
AAS74426
ID   AAS74426 standard; cDNA; 2277 BP.
XX
AC   AAS74426;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #10230.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG10239.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 10230; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2277 BP; 878 A; 483 C; 473 G; 443 T; 0 other;
```

Alignment Scores:			
Pred. No.:	0.609	Length:	2277
Score:	109.50	Matches:	73
Percent Similarity:	37.21%	Conservative:	39
Best Local Similarity:	24.25%	Mismatches:	107
Query Match:	4.64%	Indels:	82
DB:	23	Gaps:	15

US-09-807-459-2 (1-458) x AAS74426 (1-2277)

```
Qy    183 SerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAla 202
Db    868 ACTTTCTTCCCAGAATTGGAAAAAACTACTTTAAAGTTCATATGGAACCAAAAAAGAGCC 927

Qy    203 LeuLysGlnIleIleArgSerAsnLeu--ProLeuAspIleGlyThrGluHisSerValS 222
Db    928 CTC------ATCGCCAAGTCAATCCTGAGCCAAAAGAACAAAGCCAGAAGCATCACGCTA 981

Qy    222 erArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaL 242
Db    982 CCTGACTTCGAACTATACTACAAGGCTACAGTAAC------CAAAACAGCCTGGTACTGG 1035

Qy    242 euProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValA 262
Db   1036 TACCAAAA------CAGAGATATAGACCAATGGAACAGAACAGAGCCCTCAGAAATAATG 1089

Qy    262 laGlyTyr---------------Val-AspThrPro------------------------ 268
Db   1090 CCGCATATCTACAACTATCTGATCTTTGACAAACCTCACAAAAACAAGCAATGGGGAAAG 1149
```

```
Qy    269 ------------------Trp---TyrLysLysTrpTyrMetLysLeuLysAsnPheMet 281
                            |||   ::::::||||||            :::
Db   1150 GATTCCCTATTTAATAAATGGTGCTGGGAAAAATGGCTAACCATATGTAGAAAGCTGAAA 1209

Qy    282 ValAsnArgValPheIleProThrLysLysPhePheAsnLysGluIleArgGlu------ 299
            ::::::            |||      |||      ::::::   |||::::::
Db   1210 CTGGATCCCTTCCTTACACCTTATACAAAAATTAATTCAAGATGGATTAAAGACTTACAT 1269

Qy    300 ------------ProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeu 315
                       |||   |||:::||||||:::           :::|||
Db   1270 GTTAGAGCTAAAACCATAAAAACCCTAGAAGAAAAACTAGGCAATACCATTCAGGAC--- 1326

Qy    316 PheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspPro 335
                  ||||||   |||      ||||||   :::|||            |||
Db   1327 ------------ATAGGCATGGGCAAG---GACTTCATGTCTAAA-----------CCA 1359

Qy    336 SerLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLys--- 354
            ||||||         |||         |||      ||||||   :::   |||
Db   1360 CCAAAAGCAACGGCAGCAAAAGCCAAAATTGACAAATGGGATCTAATTAAACTAAAGAAC 1419

Qy    355 ------IleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLys 372
                  :::   |||         :::|||         |||   |||:::|||
Db   1420 TTCTGCACAGCAAAAGAAACTACCATCAGAGTGAAC---------AGGCAACCTACAAAA 1470

Qy    373 AlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGln 392
                                  |||   :::|||
Db   1471 --------------------------TGGGAGAAAATTTTCACAACC------------ 1491

Qy    393 GlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLys 412
                                                  |||   ||||||   :::
Db   1492 ------------------------------TACTCATCTGACAAAGGGCTAATATCCAGA 1521

Qy    413 ValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhe 432
            :::|||   |||         :::::::::   |||
Db   1522 ATCTACAATGAACTCCAGCAAATTTACAAGAAAAAA----------------------- 1557

Qy    433 IleAsnLysGluIleArgAspProSerLysAlaLeuIleArgLysValSerThrGlu 451
               |||      |||:::      :::|||   :::   |||      |||   |||
Db   1558 ACAAACAACCCCATCAAAAAGTGGGCGAAGGACATGAACAGACACTTCTCAAAAGAA 1614

RESULT 38
AAS74599
ID   AAS74599 standard; cDNA; 2277 BP.
XX
AC   AAS74599;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #10403.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG10412.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 10403; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2277 BP; 878 A; 483 C; 473 G; 443 T; 0 other;

Alignment Scores:
Pred. No.:              0.609          Length:          2277
Score:                  109.50         Matches:         73
Percent Similarity:     37.21%         Conservative:    39
Best Local Similarity:  24.25%         Mismatches:      107
Query Match:            4.64%          Indels:          82
DB:                     23             Gaps:            15

US-09-807-459-2 (1-458) x AAS74599 (1-2277)

Qy    183 SerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAla 202
            :::||||||   :::|||      ||||||      |||   |||      |||||||||
Db    868 ACTTTCTTCCCAGAATTGGAAAAAACTACTTTAAAGTTCATATGGAACCAAAAAAGAGCC 927

Qy    203 LeuLysGlnIleIleArgSerAsnLeu--ProLeuAspIleGlyThrGluHisSerValS 222
            |||      |||   :::|||   |||  |||   :::         :::|||      :
Db    928 CTC------ATCGCCAAGTCAATCCTGAGCCAAAAGAACAAAGCCAGAAGCATCACGCTA 981

Qy    222 erArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaL 242
            ::   |||:::   |||         |||   :::                  :::      :
Db    982 CCTGACTTCGAACTATACTACAAGGCTACAGTAAC------CAAAACAGCCTGGTACTGG 1035

Qy    242 euProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValA 262
            ::|||||      :::|||:::      |||         |||   |||       |
Db   1036 TACCAAAA------CAGAGATATAGACCAATGGAACAGAACAGAGCCCTCAGAAATAATG 1089

Qy    262 laGlyTyr---------------Val-AspThrPro----------------------- 268
            ||   |||               ::: |||   |||
Db   1090 CCGCATATCTACAACTATCTGATCTTTGACAAACCTCACAAAAACAAGCAATGGGGAAAG 1149

Qy    269 ------------------Trp---TyrLysLysTrpTyrMetLysLeuLysAsnPheMet 281
                            |||   ::::::||||||            :::
Db   1150 GATTCCCTATTTAATAAATGGTGCTGGGAAAAATGGCTAACCATATGTAGAAAGCTGAAA 1209

Qy    282 ValAsnArgValPheIleProThrLysLysPhePheAsnLysGluIleArgGlu------ 299
            ::::::            |||      |||      ::::::   |||::::::
Db   1210 CTGGATCCCTTCCTTACACCTTATACAAAAATTAATTCAAGATGGATTAAAGACTTACAT 1269

Qy    300 ------------ProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeu 315
                       |||   |||:::||||||:::           :::|||
Db   1270 GTTAGAGCTAAAACCATAAAAACCCTAGAAGAAAAACTAGGCAATACCATTCAGGAC--- 1326

Qy    316 PheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspPro 335
```

```
                    ||||||   |||      ||||||   :::|||            |||
Db   1327 ------------ATAGGCATGGGCAAG---GACTTCATGTCTAAA------------CCA 1359

Qy    336 SerLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLys--- 354
          ||||||         |||         |||      ||||||   :::   |||
Db   1360 CCAAAAGCAACGGCAGCAAAAGCCAAAATTGACAAATGGGATCTAATTAAACTAAAGAAC 1419

Qy    355 ------IleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLys 372
                :::   |||         :::|||         |||   |||:::|||
Db   1420 TTCTGCACAGCAAAAGAAACTACCATCAGAGTGAAC---------AGGCAACCTACAAAA 1470

Qy    373 AlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGln 392
                                       |||   :::|||
Db   1471 --------------------------TGGGAGAAAATTTTCACAACC------------ 1491

Qy    393 GlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLys 412
                                                 |||   ||||||   :::
Db   1492 ----------------------------TACTCATCTGACAAAGGGCTAATATCCAGA 1521

Qy    413 ValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhe 432
          :::|||   |||         :::::::::   |||
Db   1522 ATCTACAATGAACTCCAGCAAATTTACAAGAAAAAA----------------------- 1557

Qy    433 IleAsnLysGluIleArgAspProSerLysAlaLeuIleArgLysValSerThrGlu 451
          |||      |||:::     :::|||   :::   |||      |||   |||
Db   1558 ACAAACAACCCCATCAAAAAGTGGGCGAAGGACATGAACAGACACTTCTCAAAAGAA 1614

RESULT 39
AAS79126
ID   AAS79126 standard; cDNA; 2277 BP.
XX
AC   AAS79126;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #14930.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG14939.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 14930; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2277 BP; 878 A; 483 C; 473 G; 443 T; 0 other;

Alignment Scores:
Pred. No.:              0.609          Length:          2277
Score:                  109.50         Matches:         73
Percent Similarity:     37.21%         Conservative:    39
Best Local Similarity:  24.25%         Mismatches:      107
Query Match:            4.64%          Indels:          82
DB:                     23             Gaps:            15

US-09-807-459-2 (1-458) x AAS79126 (1-2277)

Qy    183 SerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAla 202
          :::|||||||   :::|||      ||||||      |||   |||      |||||||||
Db    868 ACTTTCTTCCCAGAATTGGAAAAAACTACTTTAAAGTTCATATGGAACCAAAAAAGAGCC 927

Qy    203 LeuLysGlnIleIleArgSerAsnLeu--ProLeuAspIleGlyThrGluHisSerValS 222
          |||      |||   :::|||   |||  |||   :::         :::|||      :
Db    928 CTC------ATCGCCAAGTCAATCCTGAGCCAAAAGAACAAAGCCAGAAGCATCACGCTA 981

Qy    222 erArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaL 242
          ::   |||:::   |||         |||   :::                :::      :
Db    982 CCTGACTTCGAACTATACTACAAGGCTACAGTAAC------CAAAACAGCCTGGTACTGG 1035

Qy    242 euProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValA 262
          ::||||||      :::|||:::      |||         |||   |||          |
Db   1036 TACCAAAA------CAGAGATATAGACCAATGGAACAGAACAGAGCCCTCAGAAATAATG 1089

Qy    262 laGlyTyr---------------Val-AspThrPro------------------------ 268
          ||   |||               ::: |||   |||
Db   1090 CCGCATATCTACAACTATCTGATCTTTGACAAACCTCACAAAAACAAGCAATGGGGAAAG 1149

Qy    269 ------------------Trp---TyrLysLysTrpTyrMetLysLeuLysAsnPheMet 281
                           |||   ::::::||||||           :::
Db   1150 GATTCCCTATTTAATAAATGGTGCTGGGAAAAATGGCTAACCATATGTAGAAAGCTGAAA 1209

Qy    282 ValAsnArgValPheIleProThrLysLysPhePheAsnLysGluIleArgGlu------ 299
          ::::::           |||      |||      ::::::   |||::::::
Db   1210 CTGGATCCCTTCCTTACACCTTATACAAAAATTAATTCAAGATGGATTAAAGACTTACAT 1269

Qy    300 ------------ProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeu 315
                       |||   |||:::||||||:::           :::|||
Db   1270 GTTAGAGCTAAAACCATAAAAACCCTAGAAGAAAAACTAGGCAATACCATTCAGGAC--- 1326

Qy    316 PheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspPro 335
                |||||   |||      ||||||   :::|||                |||
Db   1327 ------------ATAGGCATGGGCAAG---GACTTCATGTCTAAA------------CCA 1359

Qy    336 SerLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLys--- 354
          ||||||         |||         |||      ||||||   :::   |||
Db   1360 CCAAAAGCAACGGCAGCAAAAGCCAAAATTGACAAATGGGATCTAATTAAACTAAAGAAC 1419

Qy    355 ------IleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLys 372
                :::   |||         :::|||         |||   |||:::|||
Db   1420 TTCTGCACAGCAAAAGAAACTACCATCAGAGTGAAC---------AGGCAACCTACAAAA 1470

Qy    373 AlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGln 392
                                       |||   :::|||
Db   1471 --------------------------TGGGAGAAAATTTTCACAACC------------ 1491
```

```
Qy    393 GlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLys 412
                                                  |||   ||||||   :::
Db   1492 ----------------------------TACTCATCTGACAAAGGGCTAATATCCAGA 1521

Qy    413 ValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhe 432
          :::|||   |||         :::::::::   |||
Db   1522 ATCTACAATGAACTCCAGCAAATTTACAAGAAAAAA---------------------- 1557

Qy    433 IleAsnLysGluIleArgAspProSerLysAlaLeuIleArgLysValSerThrGlu 451
            |||      |||:::      :::|||   :::   |||      |||   |||
Db   1558 ACAAACAACCCCATCAAAAAGTGGGCGAAGGACATGAACAGACACTTCTCAAAAGAA 1614

RESULT 40
AAS84093
ID   AAS84093 standard; cDNA; 2394 BP.
XX
AC   AAS84093;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #19897.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG19906.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 19897; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
```

```
SQ   Sequence 2394 BP; 935 A; 504 C; 490 G; 465 T; 0 other;

Alignment Scores:
Pred. No.:              0.651          Length:          2394
Score:                  109.50         Matches:         73
Percent Similarity:     37.21%         Conservative:    39
Best Local Similarity:  24.25%         Mismatches:      107
Query Match:            4.64%          Indels:          82
DB:                     23             Gaps:            15

US-09-807-459-2 (1-458) x AAS84093 (1-2394)

Qy    183 SerPhePheAsnLysLeuSerPheThrThrGlyLeuPheGlyTrpGlyIleLysArgAla 202
          :::||||||   :::|||      ||||||      |||   |||      |||||||||
Db    985 ACTTTCTTCCCAGAATTGGAAAAAACTACTTTAAAGTTCATATGGAACCAAAAAAGAGCC 1044

Qy    203 LeuLysGlnIleIleArgSerAsnLeu--ProLeuAspIleGlyThrGluHisSerValS 222
          |||      |||   :::|||   |||  |||   :::         :::|||      :
Db   1045 CTC------ATCGCCAAGTCAATCCTGAGCCAAAAGAACAAAGCCAGAAGCATCACGCTA 1098

Qy    222 erArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaL 242
          ::   |||:::   |||         |||   :::                 :::      :
Db   1099 CCTGACTTCGAACTATACTACAAGGCTACAGTAAC------CAAAACAGCCTGGTACTGG 1152

Qy    242 euProLysPheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValA 262
          ::||||||      :::|||:::      |||         |||  |||         |
Db   1153 TACCAAAA------CAGAGATATAGACCAATGGAACAGAACAGAGCCCTCAGAAATAATG 1206

Qy    262 laGlyTyr---------------Val-AspThrPro------------------------ 268
          ||   |||               ::: |||   |||
Db   1207 CCGCATATCTACAACTATCTGATCTTTGACAAACCTCACAAAAACAAGCAATGGGGAAAG 1266

Qy    269 ------------------Trp---TyrLysLysTrpTyrMetLysLeuLysAsnPheMet 281
                           |||   ::::::||||||           :::
Db   1267 GATTCCCTATTTAATAAATGGTGCTGGGAAAAATGGCTAACCATATGTAGAAAGCTGAAA 1326

Qy    282 ValAsnArgValPheIleProThrLysLysPhePheAsnLysGluIleArgGlu------ 299
          ::::::           |||     |||     ::::::   |||:::::::
Db   1327 CTGGATCCCTTCCTTACACCTTATACAAAAATTAATTCAAGATGGATTAAAGACTTACAT 1386

Qy    300 ------------ProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeu 315
                       |||   |||:::|||||:::           :::|||
Db   1387 GTTAGAGCTAAAACCATAAAAACCCTAGAAGAAAAACTAGGCAATACCATTCAGGAC--- 1443

Qy    316 PheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspPro 335
                  ||||||   |||      ||||||   :::|||            |||
Db   1444 ------------ATAGGCATGGGCAAG---GACTTCATGTCTAAA-----------CCA 1476

Qy    336 SerLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLys--- 354
            ||||||         |||         |||      ||||||   :::   |||
Db   1477 CCAAAAGCAACGGCAGCAAAAGCCAAAATTGACAAATGGGATCTAATTAAACTAAAGAAC 1536

Qy    355 ------IleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLys 372
                  :::   |||         :::|||         |||   |||:::|||
Db   1537 TTCTGCACAGCAAAAGAAACTACCATCAGAGTGAAC---------AGGCAACCTACAAAA 1587

Qy    373 AlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGln 392
                                    |||   :::|||
Db   1588 --------------------------TGGGAGAAAATTTTCACAACC------------ 1608

Qy    393 GlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLys 412
                                                  |||   ||||||   :::
Db   1609 ----------------------------TACTCATCTGACAAAGGGCTAATATCCAGA 1638

Qy    413 ValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhe 432
          :::|||   |||         :::::::::   |||
Db   1639 ATCTACAATGAACTCCAGCAAATTTACAAGAAAAAA---------------------- 1674

Qy    433 IleAsnLysGluIleArgAspProSerLysAlaLeuIleArgLysValSerThrGlu 451
            |||      |||:::      :::|||   :::   |||      |||   |||
Db   1675 ACAAACAACCCCATCAAAAAGTGGGCGAAGGACATGAACAGACACTTCTCAAAAGAA 1731
```

Search completed: October 19, 2002, 08:20:16
Job time : 1278 secs